Graz University of Technology

Institute for Computer Graphics and Vision

Dissertation

Enhancing Handheld Navigation Systems with Augmented Reality

Alessandro Mulloni

Graz, Austria, July 2012

Thesis supervisors
Prof. Dr. Dieter Schmalstieg
Prof. Dr. Mark Billinghurst

To Madda

Abstract

The goal of this thesis is to design, implement and evaluate novel techniques for enhancing handheld navigation systems with Augmented Reality (AR). AR is an increasingly popular technology for handheld navigation systems but is rarely studied from a Human-Computer Interaction (HCI) perspective. We rather argue that advantages and limitations of AR must be clearly identified and taken into account when designing navigation systems that integrate AR.

In this thesis, we apply typical HCI methodologies to tackle this argument. We design and evaluate novel interfaces to support the integration of AR in handheld navigation systems, and evaluate all our designs with real-world users on common tasks of exploration and wayfinding. In particular, we analyse when AR enhances navigation systems and when it does not enhance them. In the latter case, we investigate how AR can be complemented by other interfaces to be more effective.

The results from our evaluations show that AR offers two intuitive interaction metaphors for accessing information: pointing at the environment and browsing paper maps. Further, the availability of physical props in AR, such as paper maps, also fosters and supports discussion between multiple users. AR fails when users need an overview of the information, or when tracking is inaccurate; however, our work shows that overlays and transitions to other interfaces can compensate for these shortcomings. Finally, our results suggest that decision points are key scenarios in which AR can enhance handheld navigation systems.

The work presented in this thesis deepens the understanding on how AR can be integrated into handheld navigation systems in an effective way. It also provides insight on the tracking technologies and evaluation methodologies necessary to successfully deploy and evaluate handheld AR navigation systems in real-world settings. Our contribution supports interface designers in making informed design decisions for handheld AR navigation systems, and it helps developers understanding the technical components necessary for implementing handheld AR navigation systems and deploying them on the field.

Kurzfassung

Diese Dissertation behandelt das Design, die Implementierung und die Evaluierung neuer Techniken zur Verbesserung von mobilen Navigationssystemen durch Augmented Reality (AR). Obwohl AR-Technologie im Bereich mobiler Navigationssysteme immer mehr an Popularität gewinnt, mangelt es an wissenschaftlicher Forschung im Zusammenhang mit Human-Computer Interaction (HCI). Bei Navigationssystemen mit integrierter AR muss vorrangig die Abwägung von Vor- und Nachteilen diskutiert werden.

Diese Arbeit behandelt dieses Thema anhand typischer HCI-Methoden. Wir entwickeln neuartige Interfaces, die die Integration von AR in mobilen Navigationssysteme ermöglichen, und evaluieren alle Designs im Rahmen von gewöhnlichen Navigationssaufgaben mit Benutzern. Insbesondere untersuchen wir, wann AR die Navigation unterstützt, und falls AR sie nicht untersützt, wie sie mit anderen Interfaces vervollständigt werden kann, um die Effizienz zu steigern.

Unsere Ergebnisse zeigen, dass AR zwei intuitive Interaktionsmetaphern für den Informationszugang bietet: das Zeigen auf die Umgebung und das Durchsuchen von Karten. Weiterhin fördert und unterstützt die Verwendung von physischen Hilfsmitteln in AR – wie etwa Landkarten – die Zusammenarbeit zwischen mehreren Benutzern. Allerdings scheitert AR, dem Benutzer einen Überblick über Informationen zu geben, oder wenn die Positionierung fehlerhaft ist. Diese Arbeit zeigt, wie die Überlagerung mit, und der Übergang zu anderen Interfaces diese Mängel kompensieren können. Die Ergebnisse zeigen, dass Entscheidungspunkte jene Schlüsselszenarien sind, bei denen AR mobile Navigationssysteme aufwerten kann.

In dieser Arbeit wird vertieft darauf eingegangen, wann AR in mobile Navigationssysteme effektiv integriert werden kann. Sie gibt Einblick in Trackingtechnologien und Evaluierungsmethoden, die für die erfolgreiche Anwendung sowie Bewertung mobiler AR-Navigationssysteme in realen Umgebungen notwendig sind. Dieser Beitrag ermöglicht es Interface-Designern, fundierte Entscheidungen für die Gestaltung von mobilen AR-Navigationssystemen zu treffen. Weiters untersützt sie Entwickler beim Verständnis der technischen Komponenten, die für die Implementierung und Anwendung von mobilen AR-Navigationssystemen in der Welt notwendig sind.

Acknowledgments

Research is also teamwork, and this thesis would not have been possible without my friends and colleagues from the ICG. First of all, I thank my advisor Prof. Dieter Schmalstieg: he introduced me into this research field, he gave precious advices and he let me freedom to pursue my research goals, also when they diverged from the research agenda of the lab. I also thank Daniel Wagner and Hartmut Seichter: they were a driving force for my research work at the CDL and they contributed a lot to this research work. I also thank all other members of the CDL: Tobias Langlotz and Lukas Gruber who started this PhD adventure with me, and all later members; they all supported, influenced and inspired this research work over the years. Thanks to Ernst Kruijff and Eduardo Veas for the discussions on navigation-related topics, and to all other colleagues at the ICG for trying out my weirdest prototypes, for tolerating all the study participants that roamed around the buildings, or even for just having a relaxing chat at the coffee machine in the kitchen.

I had the pleasure (and honour) to visit a number of research labs, and I thank all institutions that hosted me and made me feel like home. I am particularly thankful to all the brilliant researchers I encountered during these travels – I learned a lot from you guys! Thanks to Prof. Mark Billinghurst (also for the great feedback he gave on this thesis), and to all his team at the HITLab NZ in Christchurch; a special thanks to Andreas Dünser for his input on many ideas and experimental designs in this thesis. Thanks to Ann Morrison and all the people I worked with while I was at the HIIT in Helsinki. Thanks to Prof. Patrick Baudisch and all his team at the Hasso-Plattner Institute in Potsdam.

I am grateful to my parents and my sister, for all the support they gave me throughout (and outside) my studies – it's great to know that you are always supporting and endorsing me. One of the best side effects of this thesis is to know that you are proud and happy because of my accomplishments.

Finally, the biggest thanks goes to Maddalena. You gave me stability, happiness, and a good amount of motivational pushes. You also questioned and broke all the early prototypes that I put into you hands – you might have felt sorry back then, but it made my research much more solid. This thesis would have looked much thinner and weaker without your support: the least I can do is to dedicate it to you.

Contents

List of Figures

List of Tables

PART I

Overview

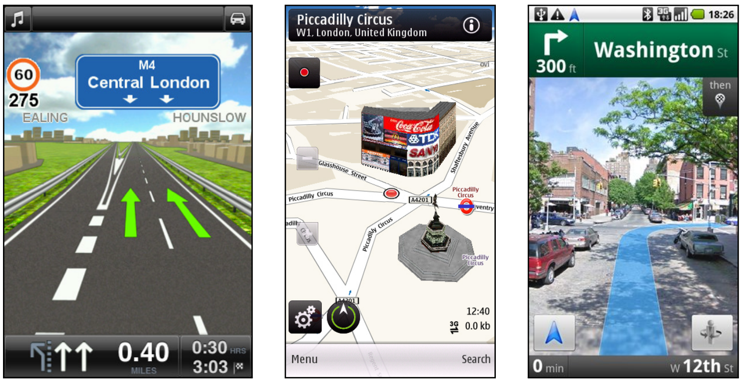
Chapter 1

Introduction

Handheld navigation systems have become widespread over the last decade, and there have been huge improvements in the computing and sensing capabilities of these devices. Similarly, in the last few years, a large amount of aerial and street-level imagery, 3D reconstruction data and geo-tagged hypermedia has enabled the development of rich interfaces for handheld navigation systems, which go way beyond simple 2D-map interfaces.

1.1 A trend towards first-person views

The trend towards first-person views is clear in the latest handheld navigation systems. Figure 1.1 shows a few examples of this trend: TomTom [192] uses first-person renderings to support car drivers; OVI Maps [130] are enhanced with 3D renderings of prominent landmarks, to give users clear visual anchors for matching the map view and the physical environment; and Google Maps Navigation [44] augments street-level imagery with a



(a) (b) (c)

Figure 1.1: Examples of first-person views in modern handheld navigation systems. (a) TomTom [192], (b) OVI Maps [130], (c) Google Maps Navigation [44].

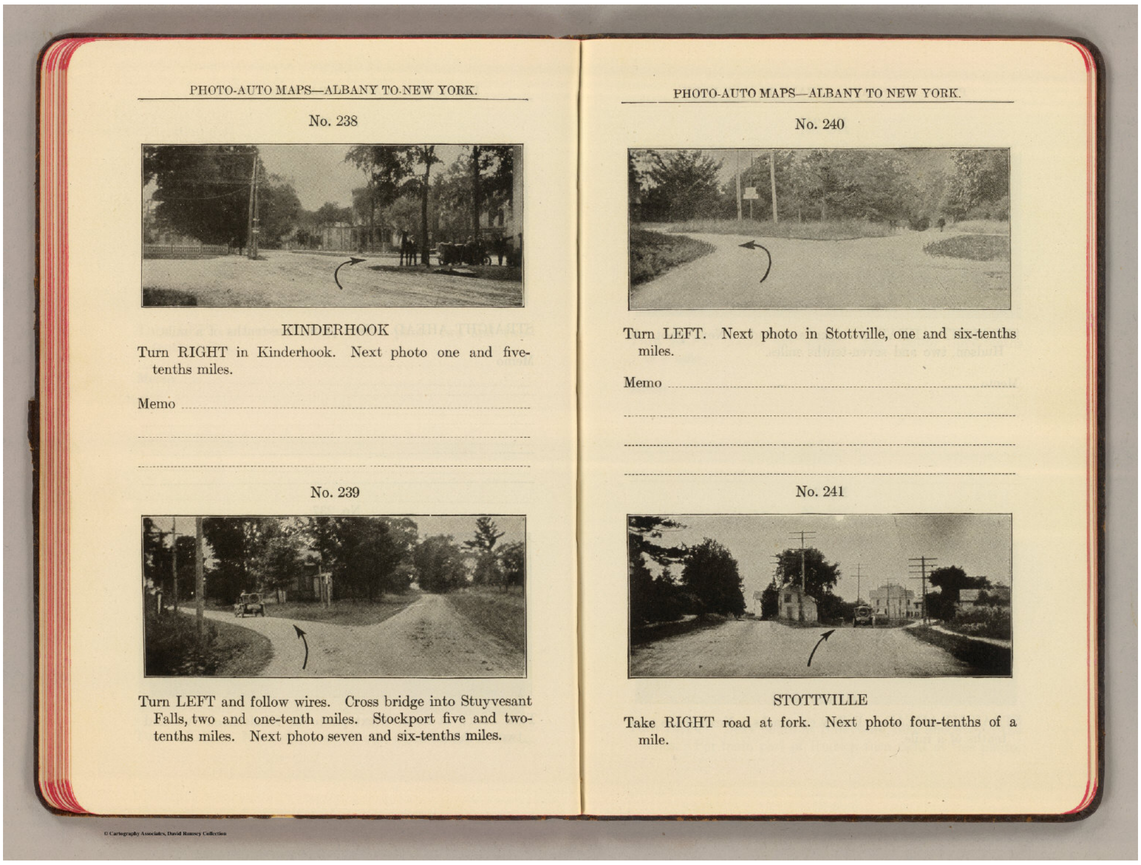


PHOTO-AUTO MAPS—ALBANY TO NEW YORK.

No. 238

KINDERHOOK

Turn RIGHT in Kinderhook. Next photo one and five-tenths miles.

Memo ....

No. 239

Turn LEFT and follow wires. Cross bridge into Stuyvesant Falls, two and one-tenth miles. Stockport five and two-tenths miles. Next photo seven and six-tenths miles.

PHOTO-AUTO MAPS—ALBANY TO NEW YORK.

No. 240

Turn LEFT. Next photo in Stottville, one and six-tenths miles.

Memo ....

No. 241

STOTTVILLE

Take RIGHT road at fork. Next photo four-tenths of a mile.

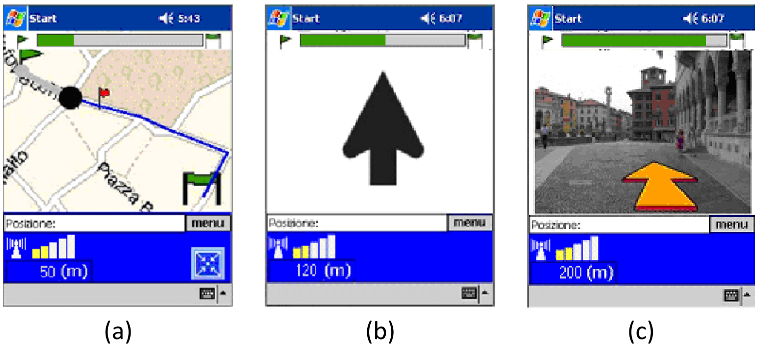
© Cartography Associates, David Rumsey Collection

Figure 1.2: An early example of augmented first-person imagery by Chapin [20]. Reprinted from `http://www.davidrumsey.com/luna/servlet/detail/RUMSEY~8~1~34014~1170167` under the Creative Commons License.

virtual path to aid users' navigation through first-person views.

Supporting navigation with first-person views is a well-known idea, already exploited by Chapin [20] in the early 1900s in his popular series *Photo-auto maps* (see Figure 1.2). The idea is rooted in the way that humans navigate: first-person knowledge (such as the visual appearance of intersections) is quickly absorbed by humans during navigation, whereas third-person knowledge (such as maps) takes longer to develop [190]. Furthermore, not all people are equally good at developing, maintaining and using third-person knowledge. Recently, various studies [24, 58, 204] supported the intuition of Chapin with empirical evidence, highlighting that enhancing handheld navigation systems with augmented photographs (see Figure 1.3) improves navigation. In particular, augmented photographs help making correct navigational decisions faster, compared to map-based interfaces.

Compared to other first-person interfaces, Augmented Reality (AR) offers one unique



(a) (b) (c)

Figure 1.3: Comparative evaluations between augmented photographs and other interfaces. Sample screenshots from a user study by Chittaro et al. [24] that compared maps (a), compass-like arrows (b) and augmented photographs (c) in an outdoor navigation scenario.

advantage: due to the use of live-video, the augmented view exactly matches the actual appearance of the environment. In contrast, pre-recorded data typically differs from the current appearance of the environment, due to a number of factors such as seasonal changes, daylight differences, moving objects, and the offset between the user and the position from which the image was originally recorded. Intuitively, AR should be able to provide better support for navigation than any other interface based on pre-recorded data.

1.2 Problem statement

A large body of previous work (discussed in the following chapter) effectively applies AR to various navigation scenarios, including exploratory and wayfinding tasks, both indoors and outdoors.

However, AR has a number of shortcomings that impacts its usability in navigation scenarios. First, the information is augmented only within the field of view and from the viewpoint of the video camera. This limits the overview achievable and the possible viewpoints on it. Second, AR requires continuous tracking of the position and orientation (the *pose*) of the user's device. Tracking is not always achievable with the necessary accuracy, because of current technological limitations (e.g., sensor accuracy) and algorithmic shortcomings – for example vision-based tracking in unknown environments is a known hard, and not fully solved, problem. Finally, the necessity to continuously hold the handheld device upright, in order to point the camera at the environment, poses ergonomic and social questions: the user may not be willing to continuously point the device at her surroundings, and that behaviour may not be socially acceptable.

Previous work applies AR to navigation mostly with a technical focus, as an application scenario to showcase or evaluate technological improvements. As the focus is principally on advancing the underlying technology, usability issues are only marginally considered. There is a general lack of work that takes an Human-Computer Interaction (HCI) perspective on the usability of AR interfaces to support navigation tasks and there is little experimental evidence to support the argument that *AR enhances human performance* in such tasks.

1.3 Research questions

This thesis takes an HCI perspective on handheld AR navigation systems, targeting research questions that have not been thoroughly addressed yet:

Q1. When* does AR enhance handheld navigation systems?

Q2. When does it fail to enhance them?

Q3. In the latter cases, how can we complement AR with other interfaces to build a hybrid and more effective interface?

Our basic research hypothesis is that *AR can enhance handheld navigation systems only (1) if its advantages and limitations, for each specific task, are clearly identified and (2) taken into account in the interface design, and (3) if the limitations are properly addressed by supporting AR with complementary interfaces.*

In this thesis, we test this hypothesis through standard HCI methodologies: we first select representative tasks and scenarios for handheld AR navigation systems, and design interfaces to support the selected tasks. We then conduct user evaluations to gain qualitative and quantitative insight on the usability of our interfaces, and further understanding on the conditions in which AR can effectively enhance handheld navigation systems. We finally use the results from our evaluations to reflect on our interface designs, to propose corrections to them, and in some cases to re-iterate the design process.

1.4 Contributions

This thesis' contribution is focused on two main navigational tasks: **exploration** and **wayfinding**. Exploration refers to users exploring the space that surrounds them.

*As typical in HCI, "when" denotes a combination of *user population*, *task* and all other *contextual conditions* related to specific experiments.

Wayfinding refers to users physically moving in space to reach a chosen destination. In everyday life, these two tasks are complementary and typically intermixed – for example, we can think of a user exploring the restaurants in her surroundings, choosing the most interesting one, and finally wayfinding to get to it. However, to clearly separate the impact of AR on each of the two tasks, we avoid overly complex experimental designs and study them separately. For both exploration and wayfinding, this thesis contributes a number of novel interface designs, and the results and insight from several user evaluations of such designs.

For exploration, we consider two possible ways of providing AR support. First, we consider using AR to give an **egocentric view** on the information – in this case, information is augmented directly on its corresponding physical anchor in the environment (for example, a shop's name is augmented on the shop's window). We discuss the case in which the interface *knows that the user is looking for a specific piece of information* and must help the user *understand its location in the environment.* We contribute a comparative study between AR and other state-of-the-art non-AR interfaces, that gives understanding on the advantages and shortcomings of AR for this task. We then discuss the more generic case in which the system *does not know exactly what the user is looking for* and the interface must *provide a full overview on all surrounding information.* We contribute two novel zooming-interface designs that support this overview, and evaluate their usability with a pilot study and two comparative studies with a state-of-the-art non-AR interface. We conduct a further design iteration on one of the two zooming interfaces, contributing a novel interface design for browsing panoramic overviews, applicable not only to AR but also to generic location-based systems. We evaluate its usability with one pilot study and three controlled studies. Finally, we consider using AR to give an **exocentric view** on the information, as a complementary support to egocentric views – in this case, information is not augmented on the environment but on its corresponding location on a paper map. We contribute an exploratory user study to gain insight on how users operate augmented maps in real-world tasks.

For wayfinding, we first consider supporting **outdoor navigation** with AR. We contribute a preliminary study that investigates how often, where, and how AR is useful in such scenario. We then consider **indoor navigation** as a more challenging scenario, because of the complex three-dimensional paths and the difficulty of continuously tracking the user's position. We contribute three iterations of interface design: in a first iteration, we study *sparse localisation* (the user's position is tracked only at certain *info points* in the building, rather than continuously) – a concept that strongly reduces tracking-infrastructure requirements for indoor navigation systems. We evaluate its validity with one controlled study and one real-world exploratory study. In the next iteration, we enhance sparse localisation with AR cues and evaluate their usability with one user study. Finally, in the last iteration we propose a novel design that transitions between AR and Virtual Reality (VR) interfaces, exploiting human navigation abilities to provide continuous navigational support in our sparse localisation scenario. We evaluate this last iteration with one further study.

The results presented in this thesis deepen the understanding on how AR can be effectively integrated into handheld navigation systems. This contribution can support interface designers in making informed design decisions. Furthermore, the large number of user evaluations presented in this thesis provides an overview of possible methodologies for evaluating handheld AR navigation systems in controlled and uncontrolled scenarios. Finally, the contribution of this thesis also has a strong technical component, as all presented interface designs were implemented and evaluated with users in realistic settings. The author made not only minor contributions to the underlying tracking technology, but also major contributions in bridging it with rapid prototyping and HCI. Overall, this thesis gives an outline of the interdisciplinarity necessary to design and develop handheld AR navigation systems, and to evaluate them under real-world conditions.

1.5 Selected publications

The following list of selected peer-reviewed publications gives an overview of the scientific activities and the collaborations which occurred during the work of this thesis.

First, the related work in this thesis is partially based on the following book chapter.

- Raphael Grasset, **Alessandro Mulloni**, Mark Billinghurst, and Dieter Schmalstieg. *Navigation Techniques in Augmented and Mixed Reality: Crossing the Virtuality Continuum.* In Borko Furht, editor, Handbook of Augmented Reality, pages 379–407. Springer New York, New York, NY, 2011. [46] [**Chapter 2: Related work**]

  The author mainly contributed the sections on human navigation and AR as a primary source of spatial information, whereas ***Raphael Grasset*** *did most of the work for the other sections of the chapter.*

The following publications are focused on tracking technology. The author was mostly involved in the discussion of the algorithms, making minimal contributions to the implementation and optimisation of the actual trackers, but major contributions were made to the final evaluations of the tracking accuracy. These papers are tightly connected to this thesis as they form the enabling technology for all interface designs: the robust and accurate tracking algorithms are the foundations that allowed the implementation of all prototypes, and their evaluation under real-world conditions.

- Daniel Wagner, Gerhard Reitmayr, **Alessandro Mulloni**, Tom Drummond, and Dieter Schmalstieg. *Pose tracking from natural features on mobile phones.* In 2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, pages 125–134, Cambridge, UK, September 2008. [201] [**Chapter 3: Tracking planar targets**]

- Daniel Wagner, Gerhard Reitmayr, **Alessandro Mulloni**, Tom Drummond, and Dieter Schmalstieg. *Real-Time Detection and Tracking for Augmented Reality on Mobile Phones.* IEEE Transactions on Visualization and Computer Graphics, 16(3):355–368, May 2010. [202] [**Chapter 3: Tracking planar targets**]

  *For both papers, the author made minor contributions to the idea and implementation of the SIFT-based tracker, which was principally conceived and developed by **Daniel Wagner**. **Gerhard Reitmayr** did most of the work on the Ferns-based tracker. The author contributed mostly to the implementation and execution of the performance tests, and the analysis and presentation of the results.*

- Daniel Wagner, **Alessandro Mulloni**, Tobias Langlotz, and Dieter Schmalstieg. *Real-time panoramic mapping and tracking on mobile phones.* In 2010 IEEE Virtual Reality Conference (VR), pages 211–218, Boston, MA, USA, March 2010. [200] [**Chapter 4: Tracking panoramas**]

  *The author made minor contributions to the idea and implementation of the panorama tracker, which was principally conceived and developed by **Daniel Wagner**. The author and **Tobias Langlotz** were the main contributors of the final application-oriented part of the paper.*

- Gerhard Schall, **Alessandro Mulloni**, and Gerhard Reitmayr. *North-centred orientation tracking on mobile phones.* In ISMAR 2010, pages 267–268, Seoul, Korea (South), October 2010. [173] [**Section 4.2: World-aligned panorama tracker**]

*The author made minor contributions to the idea of the north-aligned panorama tracker, which was principally conceived by **Gerhard Reitmayr**. The author contributed the implementation on a mobile phone, while **Gerhard Schall** mainly worked on the tablet implementation. The author and **Gerhard Schall** worked jointly on the execution of the performance tests and the evaluation of the results.*

Finally, the following publications have a strong HCI focus and form the main contribution of this thesis. The author was a main contributor at all stages of the work – forming the design ideas, designing and implementing the interfaces, designing and conducting user evaluations, analysing and discussing the final results.

- Eduardo Veas, **Alessandro Mulloni**, Ernst Kruijff, Holger Regenbrecht, and Dieter Schmalstieg. *Techniques for view transition in multi-camera outdoor environments.* In Proceedings of Graphics Interface 2010, GI '10, page 193–200, Toronto, Ont., Canada, Canada, 2010. Canadian Information Processing Society. [198] [**Section 5.1: Pointing to a specific augmentation**]

  *The author was a main contributor of the design and the implementation of the transitional interface, while **Eduardo Veas** contributed the other two interfaces. All paper co-authors jointly worked on the design and the execution of the experiment, and the analysis of the results.*

- **Alessandro Mulloni**, Andreas Dünser, and Dieter Schmalstieg. *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 161–170, New York, NY, USA, September 2010. ACM. [119] [**Section 5.2: Providing overview on all augmentations**]

  *The author contributed the design and implementation of all interfaces and the pilot study, while he jointly worked with **Andreas Dünser** on the design and execution of the user evaluations.*

- **Alessandro Mulloni**, Hartmut Seichter, Andreas Dünser, Patrick Baudisch, and Dieter Schmalstieg. *360° Panoramic Overviews for Location-Based Services.* In Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems, CHI '12, page 2565–2568, New York, NY, USA, 2012. ACM. [120] [**Section 5.3: A digression on the design space of panoramic overviews**]

*The idea of exploring the design space of panoramic overviews started from a discussion between the author and **Patrick Baudisch**. The author worked jointly with **Hartmut Seichter** and **Patrick Baudisch** on designing the visualisation techniques, and designing and executing the evaluations. The author also contributed the implementation of all prototypes. **Andreas Dünser** collaborated in designing and executing the experiments.*

- Ann Morrison, **Alessandro Mulloni**, Saija Lemmelä, Antti Oulasvirta, Giulio Jacucci, Peter Peltonen, Dieter Schmalstieg, and Holger Regenbrecht. *Collaborative use of mobile augmented reality with paper maps.* Computers & Graphics, 35(4):789–799, August 2011. [117] [**Chapter 6: Exocentric exploration**]

  *The author was not involved in the first design iteration of the prototype, which was previous work by **Ann Morrison** et al. [118]. The author was involved in the design and implementation of the second iteration of the prototype, although a large part of the work is based on a previous implementation by **Antti Juustila** from the University of Oulu, Finland. All paper co-authors jointly worked on the design and execution of the evaluation of the second prototype.*

- **Alessandro Mulloni**, Hartmut Seichter, and Dieter Schmalstieg. *Enhancing Handheld Navigation Systems with Augmented Reality.* In Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, Workshop on Mobile Augmented Reality, 2011. [121] [**Chapter 7: Outdoor navigation**]

- **Alessandro Mulloni**, Hartmut Seichter, and Dieter Schmalstieg. *User experiences with augmented reality aided navigation on phones.* In Proceedings of ISMAR 2011 (Poster), pages 229–230. IEEE, October 2011. [123] [**Chapter 7: Outdoor navigation**]

  *For both papers, the author contributed the design and implementation of the prototype, and the design and execution of the evaluation. **Hartmut Seichter** collaborated with the author both on the interface design and on the evaluation.*

- **Alessandro Mulloni**, Daniel Wagner, Istvan Barakonyi, and Dieter Schmalstieg. *Indoor Positioning and Navigation with Camera Phones.* IEEE Pervasive Computing, 8(2):22–31, April 2009. [125] [**Section 8.1: Sparse localisation**]

The author made minor contributions to the interface design and the implementation of the application used in the experiments, which is heavily based on previous work by ***Istvan Barakonyi*** *and* ***Daniel Wagner****. The author contributed the design and execution of the controlled experiment and the real-world experiment at* TechReady7*, whereas all other real-world experiments were conducted by* ***Istvan Barakonyi*** *and* ***Daniel Wagner****.*

- **Alessandro Mulloni**, Hartmut Seichter, and Dieter Schmalstieg. *Handheld augmented reality indoor navigation with activity-based instructions.* In Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, MobileHCI '11, page 211–220, New York, NY, USA, 2011. ACM. [122] [**Section 8.2: Sparse localisation and Augmented-Reality cues**]

- **Alessandro Mulloni**, Hartmut Seichter, and Dieter Schmalstieg. *Indoor Navigation with Mixed-Reality World-in-Miniature Views.* In Proceedings of AVI 2012, 2012. [124] [**Section 8.3: Sparse localisation and Mixed-Reality Cues**]

  For both papers, the author contributed the design and implementation of the prototype, and the design and execution of the evaluation. ***Hartmut Seichter*** *collaborated with the author both on the interface design and on the evaluation.*

1.6 Other collaborations

The co-authors in the list of selected publications give an indication of all the people who contributed to the various pieces of work that form this thesis. However, the following people have to be explicitly mentioned, as their contribution was not limited to the scope of a single publication but had a broader impact on the work presented in this thesis.

- **Daniel Wagner** and later **Hartmut Seichter** were a driving force for the research work at the Christian Doppler Laboratory for Handheld Augmented Reality. They contributed to this thesis through countless discussions on use cases for AR in handheld navigation systems, possible interface designs, and implementation details. **Tobias Langlotz**, **Lukas Gruber** and all later members of the Christian Doppler Laboratory for Handheld AR were also involved in discussions on the research work presented in this thesis. They also contributed in generating a shared framework for mobile AR (Studierstube ES), which was of enormous help for this thesis as it allowed quick multi-platform prototyping of all interfaces presented. All prototypes presented in this thesis are developed on top of our Studierstube ES framework.

- **Ernst Kruijff** and **Eduardo Veas** shared with the author a strong research interest on the possible applications of AR as a support to human navigation. Besides our joint publication, our numerous discussions and brainstorming sessions had a broader influence in forming the ideas presented in this thesis.
- **Andreas Dünser** provided valuable suggestions on most of the experimental designs and data analysis procedures employed in this thesis.

Chapter 2

Related work

The vast amount of geo-referenced information available enables the development of rich interfaces for handheld navigation systems. Modern handheld navigation systems go beyond the first research prototypes of the past decades, which were typically based on 2D maps. Navigation systems now include photographs, 3D renderings, audio and multimodal feedback. Furthermore, interface designers of handheld navigation systems often exploit psychological knowledge of how humans navigate to design more effective and usable systems.

In the first part of this chapter, we introduce the reader to these topics: how humans form navigational goals, how they build mental models of the surrounding environment, and which environmental cues they exploit for navigation. We also show how this knowledge can be used to inform interface design.

We then discuss related work on handheld navigation systems. We focus in particular on recent trends that are related to AR and, consequently, to the topic of this thesis.

In the final part of the chapter, we present a broad discussion on related work in AR navigation systems. We discuss in particular: how AR has been integrated into the interface designs of various navigation systems, how it has often been combined with other types of interfaces, and how previous work has addressed the usability of AR navigation systems. As we discuss the results of previous studies, we also present the usability questions that are still open: these questions are the starting point of our research work, and they will lead us directly to the main contribution of this thesis. We conclude the chapter with a brief overview of the tracking technology which is necessary to develop handheld AR navigation systems.

2.1 Human navigation

Navigation is the task of moving within and around an environment, and can be divided into *travel* and *wayfinding* activities [16]. With travel, we refer to the motor component of navigation – a person performing low-level motor activities in order to control her position and orientation within the environment. With wayfinding, we refer to the cognitive component of navigation – a person understanding her position within the environment and planning a path from her current position to a chosen destination. Besides wayfinding, people typically also perform *exploratory tasks* such as browsing the environment and obtaining information about the surrounding buildings and objects. Clearly, the amount of wayfinding and exploration involved during navigation is a function of the person's goals: a person in a hurry will most likely perform exclusively goal-directed wayfinding, whereas exploration will be prominent for a tourist navigating an unknown city.

Landmarks are the foundation of human navigation, in particular for pedestrians, who are the focus of this thesis. Lynch [101] bridges urban planning and environmental psychology in a broad study on how people create mental maps of the city they live in. From a large number of participants, Lynch builds a classification of five key elements that form a mental map of a city: landmarks, paths (or routes), nodes, districts and edges. Landmarks are found to be the fixed reference points, external to the user, which can be either distant prominent elements or small local details, but always possess a strong singularity. People use landmarks as clues for the whole structure of the environment. Paths are channels through which people can travel, and all other elements of the environment are structured along and in relation to the paths. Nodes – typically the convergence of paths – are points where wayfinding decisions are often made and also have a strategic role in the environment. Edges and districts typically define borders and areas of common identity. More recent studies by Michon et al. [111] and Hirtle et al. [59] stress the value of landmarks as anchors for the navigational instructions used by people. More specifically, May et al. [106] identify two crucial roles of landmarks: the first role of landmarks is at nodes in a path, to support navigational decisions (e.g., "turn left after the church"); the second, but almost equally important, role of landmarks is along paths, to provide confirmation (e.g., "if you pass in front of the bakery, you know you are on track"). We illustrate this dual role of landmarks in Figure 2.1.

Navigation always requires the acquisition and use of spatial knowledge, and ways to structure, store and update such knowledge into a mental map [16, 28]. Spatial knowledge is typically acquired from various sources – Darken and Peterson [28] make a distinction

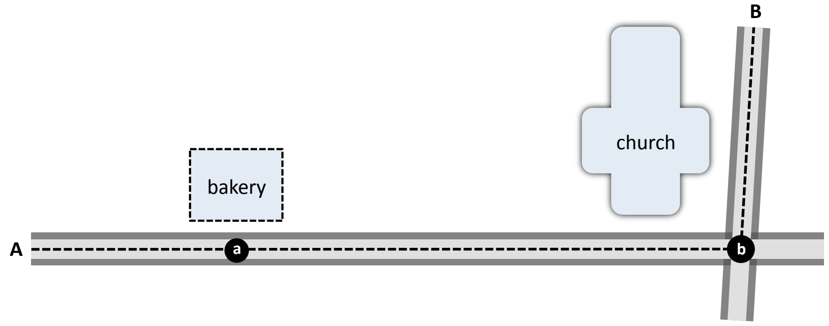


Figure 2.1: Role of landmarks in human navigation. On a path from A to B, landmarks can act both as confirmation ((a), ``if you pass in front of the bakery, you know you are on track") and as a support for navigational decisions ((b), ``turn left after the church").

between primary and secondary sources. A primary source of spatial information is the environment itself: as we navigate the environment, we extract information from it, which we can then use to support our navigational tasks. Secondary sources of spatial information are all other sources, such as a map. In the case of a user who acquires information from a secondary source, we can make a further distinction between when she is immersed in the environment related to the information (e.g., browsing a map of the surroundings) and when she is not (e.g., browsing a map while in a hotel room).

There is still no definitive model to detail how spatial knowledge is structured into a mental map. The most established and long-standing model is the Landmark, Route and Survey (LRS) model of Seigel and White [180], which was later refined by Goldin and Thorndyke [43]. The LRS model not only defines a classification of spatial knowledge, but also describes the sources from which the different classes of information can be acquired. *Landmark knowledge* represents the visual appearance of prominent cues and objects in the environment. This is the first knowledge a person develops, by directly experiencing the environment during navigation (but also through indirect exposure to it, for example by looking at photographs or videos). *Route knowledge* (or *procedural knowledge*) follows landmark knowledge and represents a point-by-point sequence of actions needed to travel a specific route. It provides information on the distance along the route, the turns and actions to be taken at each point in the sequence, and how landmarks are ordered throughout the path. Route knowledge can be acquired by navigating the route. Finally, *survey knowledge* represents the relationships between landmarks and routes in the environment in a global coordinate system. Survey knowledge typically develops over time, by numerous and repeated exposures to the environment. Browsing a map can be a shortcut for rapid acquisition of survey knowledge, but the knowledge acquired from a map tends to be orientation specific and lacks the landmark and route components, as discussed by Thorndyke et al. [190]. This type of knowledge is inferior to the knowledge obtained from repeated route traversals: for example, it causes higher errors when users are required to estimate the distance of the route from a point to another [190].

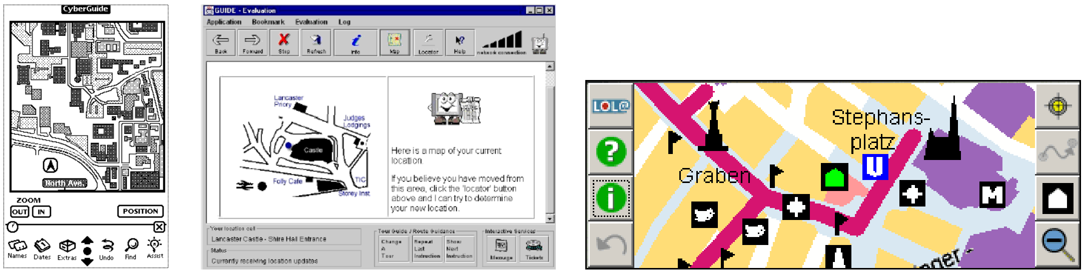
These navigation models directly inform the design of handheld navigation systems, showing not only the type of information needed by users during navigation, but also what information is needed for different types of navigational tasks. For example, route knowledge supports egocentric tasks (such as estimating orientations and route distances) better than survey knowledge, whereas survey knowledge better supports exocentric tasks (such as estimating Euclidean distances, or the relative position of generic points in the environment) [190]. Overall, people performing navigational tasks need to use various types of spatial knowledge, and multiple frames of reference. One key element of a navigation system is therefore providing support for both types of knowledge: while maps are good at supporting survey knowledge, integrating them with first person views guarantees better support for tasks that require route knowledge, and stressing the presence of certain landmarks can enhance navigational performance.

In the next section, we present an overview of previous work in handheld navigation systems. As we will show, handheld navigation systems are increasingly adding first-person views and landmarks into their interface designs, to support users with egocentric cues during their navigation.

2.2 Handheld navigation systems

Digital maps play a central role in handheld navigation systems, reflecting the fact that for centuries paper maps are the most established tool for human navigation. Cyberguide by Abowd et al. [1] and GUIDE by Davies et al. [23] pioneered the field in the 1990s; a vast number of other map-based handheld navigation systems followed later (see Figure 2.2). A broad discussion of map-based navigation systems is outside the scope of this thesis. For an introduction we refer the reader to the survey by Chen et al. [22], and the more recent work of Kenteris et al. [76]. The survey by Baus et al. [10] also gives a good introductory overview on the topic, and is part of a whole book [108] devoted to technological and user-interface topics in map-based navigation systems.

In this section, we discuss three recent trends in handheld navigation systems, which are tightly connected to our research work on handheld AR for navigation. The first topic is the integration of *first-person views* into navigation systems, and is clearly connected to



Here is a map of your current location.

If you believe you have moved from this area, click the 'locator' button above and I can try to determine your new location.

Lancaster Castle - Shire Hall Entrance

Currently receiving location updates

(a) (b) (c)

Figure 2.2: Map-based handheld navigation systems. (a) Cyberguide by Abowd et al. [1]. (b) GUIDE by Davies et al. [23]. (c) LoL@ by Pospischil et al. [146].

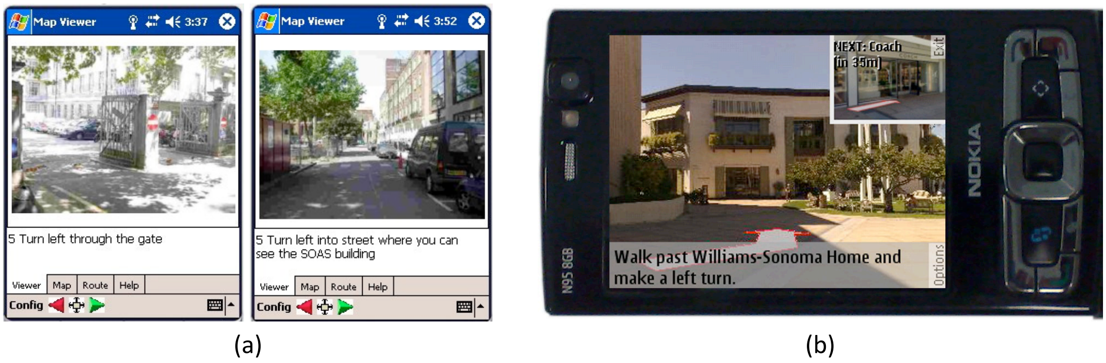
the way AR interfaces present information. The second topic is *non-visual and multimodal interaction*, which informs AR interface designers how non-visual senses should also be considered in the interface designs. The final topic is *exploiting human navigation abilities* to cope with technological limitations, and highlights how AR, which is very technologically demanding, could benefit from human abilities when the technology fails. We present and discuss the work related to each of these three topics separately, also highlighting how the related work impacts the work presented in this thesis.

2.2.1 First-person views

Handheld navigation systems are increasingly enriching map interfaces with first-person views and landmark-based cues, exploiting the egocentric quality of landmarks to support the users' understanding of the information. As we discussed in Section 2.1, this is rooted in the way humans navigate, which is often landmark-centred and based on egocentric knowledge, in particular when navigating through unknown places.

The use of landmarks in handheld navigation systems is strongly supported by related work. Millonig et al. [113] discuss the role of landmarks both as a support for navigational decisions and as checkpoints for confirmation of being on track. Their work is principally theoretical and focused on the information requirements for a pedestrian navigation system, but it sheds light on why anchoring instructions to landmarks, which are clearly and quickly understood by users, makes navigation systems more effective. The presence of landmark information is found to enhance navigation not only in map-based navigation systems (Puikkonen et al. [147], Raubal et al. [150]), but also in systems based on textual instructions (Chung et al. [27]) as well as in those based on audio instructions (Rehrl et al. [151]). However, given the focus of our research on AR, we believe that the work most related to our research are those that consider photographs or 3D renderings as enhancements for handheld navigation systems, and we discuss them in detail in the following.

Beeharee et al. [11] enhance landmark-centred textual instructions (for example, "turn left through the gate") with photographs of the corresponding landmarks (e.g., a picture of the correct gate), as shown in Figure 2.3 (a). They validate their photo-based system in a real-world user study, whose results confirm the theory of Millonig et al. [113]: photographs enhance navigation both during decision tasks and when confirmation of being on track is needed. Au et al. [5] propose using live videos from street cameras, instead of static photographs, but they do not conduct a formal evaluation of the concept. Kolbe [82] further explores the concept of photo-based navigation systems, proposing the use of *augmented photographs* – static photographs overlaid with virtual navigational cues, a concept very close to AR. His work, however, does not include a user evaluation and therefore does not provide understanding of the effectiveness of augmented photographs during navigation. Hile et al. [58] conduct an evaluation of a navigation system based on augmented photographs (Figure 2.3 (b)). A key finding from their study is that users typically use photographs only for a fraction of the overall navigation time, at decision points on the path, while they mostly rely on the map interface for the remaining part of the path. These results are consistent with those from Walther-Franks et al. [204] and Chittaro et al. [24], who also found that photographs are not used continuously during navigation, but mostly at decision points, where they improve navigational performance. All three works also provide consistent results on an intrinsic drawback of pre-recorded photographs: the photographs do not always match the actual appearance of the environment, due to seasonal changes or because they are taken from a different viewpoint than the user's.



5 Turn left through the gate

5 Turn left into street where you can see the SOAS building

Walk past Williams-Sonoma Home and make a left turn.

Figure 2.3: Use of first-person views in handheld navigation systems. (a) Beeharee et al. [11] explore the use of photographs to support text-based instructions. (b) Hile et al. [58] extend this concept further, exploring the usage of augmented photographs (photographs overlaid with virtual cues).

Wither et al. [210] investigate this issue and show that the time needed by a user to identify a landmark in a photograph progressively increases the further she is from the actual location from which the photograph was taken. Hile et al. [56, 57] try to address this issue in their most recent research, using computer vision to correctly align the augmentations and the photograph with the actual current view from the user's position.

Google Maps Mobile [44] uses street-level images to aid user navigation, often allowing users to browse them through a magic lens metaphor. The magic lens metaphor is where a handheld device acts like a lens through which the user can see an enhanced view of the real world. The user can physically turn the device at different directions in the environment in order to see corresponding spatial information. In the case of street-level images, the navigation system shows the portion of the panorama that matches the current view direction of the device. This is clearly an interaction concept very close to AR. In a recent work, Rohs et al. [163, 164] formalise magic-lens interaction, proposing a two-phase Fitts' law adaptation that better fits this novel interaction metaphor. Hürst et al. [65, 66] investigate browsing street-level panoramas with a magic lens, in real-world settings. As they discuss, users like the magic lens interaction metaphor when they can stand up and rotate freely, but they favour manual panning of the panorama with touch-screen gestures if they are more restricted in their movements (for example, sitting on a chair). This has an impact on AR, as it shows that AR might not be the best interaction choice for users who are not free to stand up and turn around. Continuously holding the device upwards in the environment also raises social concerns: not all passersby might be comfortable with this, as it might look that the user is taking a picture or a movie of them. It also raises ergonomic concerns due to the strain of holding such posture for long periods of time. These issues have not yet been studied but may have a clear impact on all magic-lens interfaces, including AR.

Zheng et al. [214] propose combining multiple street-level panoramas to give a first-person view of all the landmarks occurring during the navigation of a path. In their work, they stitch together all panoramas throughout a path, to form a continuous multi-view panorama that represents the whole path. A similar concept is implemented by Kopf et al. [84] in Street Slide, a technique that stitches all building façades of a street into a single multi-view panorama. Results from a user study are encouraging, as they show that the system is found easier to use than jumps between disjointed street-level panoramas. In their work, Kopf et al. show how Street Slide can be easily integrated into Bing Maps Mobile [83] (see Figure 2.4).

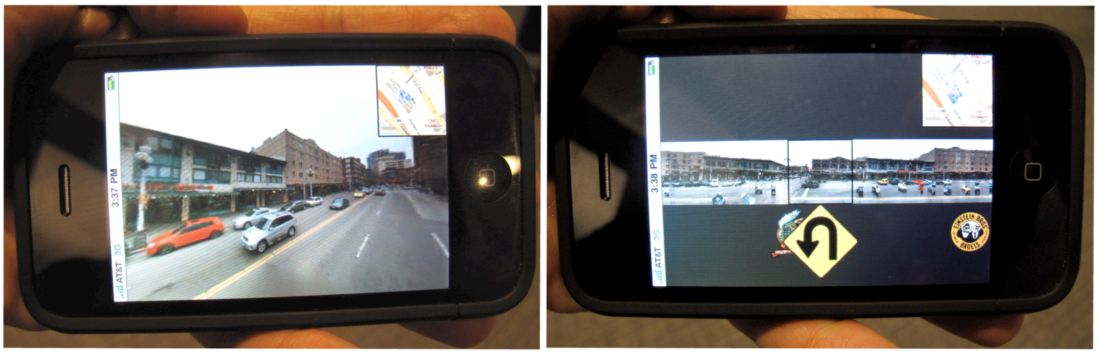


Figure 2.4: Multiview panoramas create first-person views of a long path. Kopf et al. [84] present Street-Slide, a system that allows moving between typical street-level images (left) and multi-view panoramas (right). The latter give a first-person overview of the street façades of a whole street.

Kray et al. [86] discuss using 3D maps (renderings of 3D models of the environment) in place of photographs. The results of their evaluation show that users prefer bird's eye views to street-level views, because of the higher amount of overview available when the point of view is higher above the street. The results are consistent with the findings from Fröhlich et al. [38] and Oulasvirta et al. [133] (Figure 2.5 (a)), who also find that the street-level perspective does not provide sufficient overview compared to a bird's eye view. Oulasvirta et al. also show the switch in strategy between 2D and 3D maps: 3D map users base their navigation strongly on landmarks, whereas 2D map users exploit more symbolic cues, like street names. The use of 3D maps is also validated indoors by Chittaro et al. [25, 26], who apply it to an evacuation scenario within a public building (Figure 2.5 (b)).

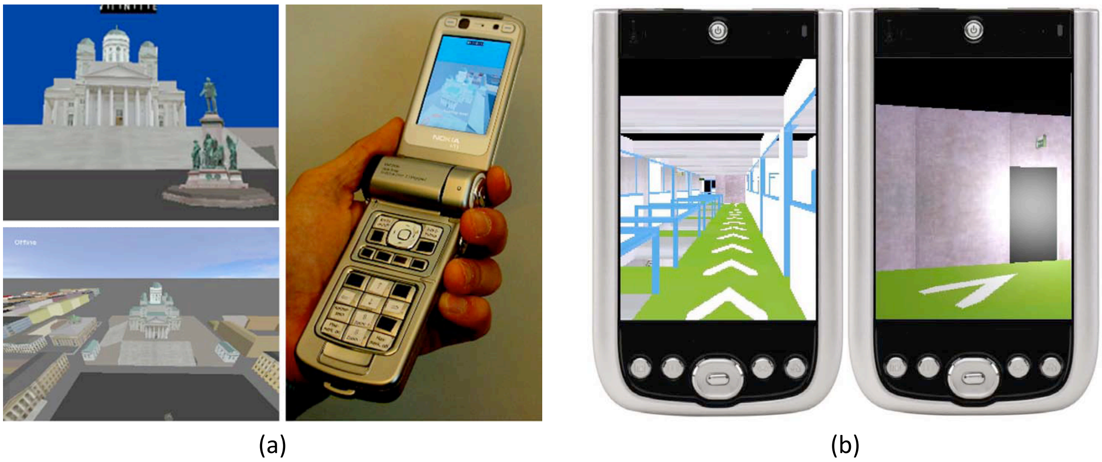


Figure 2.5: Use of 3D maps in handheld navigation systems. (a) Oulasvirta et al. [133] study the use of 3D maps in outdoor settings, discussing differences in usage between street-level views and bird's eye views. (b) Chittaro et al. [25] explore the use of 3D maps indoors, as a support for building evacuation.

This related work clearly shows the value of enhancing handheld navigation systems with first-person views, and therefore highlights the potential of adding AR views to such systems – this is a basic assumption for the whole thesis. However, previous work on the usability of AR navigation systems (presented later in this chapter) does not investigate in depth how AR can effectively improve navigation performance. For example, from the work on augmented photographs, we see that first-person views are mostly useful at decision points over a path, and users only access them for a small fraction of the overall navigation time. We use these results as a starting point for our research presented in Chapter 7, where we study how often and where users need to access AR cues during outdoor navigation, and where they rather prefer to rely on maps and audio instructions. The related work also shows that users often need more overview than the one provided by a street-level view: in Chapter 5, we study when AR does not provide sufficient overview, and how to combine it with other interfaces to provide the missing overview.

2.2.2 Non-visual and multimodal interaction

Research on handheld navigation systems is increasingly looking at how to offload the visual sense by adopting non-visual and multimodal input and output solutions. Such research originates from the need to reduce the attentional resources necessary for operating a navigation system while physically moving in the environment – for example, a recent report by Madden et al. [103] shows that one out of six adults bumped at least once onto another person, while engaged in mobile-phone usage.

Audio is often used for providing navigational instructions in a non-visual mode. For example, Rehrl et al. [151] show that timely triggered spoken instructions can suffice in guiding users, with no need for visual instructions. Other research shows that audio instructions can be embedded in a handheld navigation system in a more subtle way than spoken instructions, for example by modulating artificial sounds, as done by Holland et al. [60], or even by adapting panning and volume in music playback, as done by Jones et al. [71]. Audio can also be used as an input modality, as done by Wasinger et al. [205].

Vibration can also be used to communicate directions. For example, Rümelin et al. [168], Pielot et al. [142] (Figure 2.6 (b)) and Robinson et al. [161] show that users can successfully navigate with only the help of vibration patterns on the phone, although Pielot et al. show that users prefer to receive also visual feedback, when they are uncertain. While on-phone vibration still requires constantly holding the device in one hand, Van Erp et al. [197] explore hands-free vibrotactile waist belts and show their effectiveness

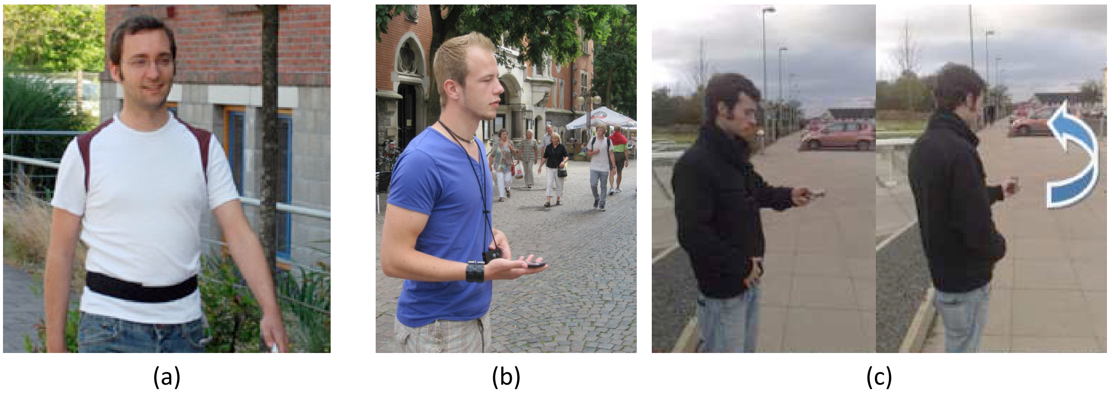


Figure 2.6: Multimodal interaction with handheld navigation systems. (a) Adding a vibro-tactile belt to communicate the direction of waypoints in the environment (Pielot et al. [141]). (b-c) Pointing as an input metaphor to explore the information in the surroundings (Pielot et al. [142], Lei et al. [93]).

in communicating directions. Pielot et al. [140, 141] show that vibrotactile belts (Figure 2.6 (a)) successfully offload users' attentive resources during navigation, but navigational performance is worse than with a map. Similarly to the phone experiments, this suggests that receiving also visual feedback is important. Vibration feedback has also been explored in contexts where attentional resources are even more precious, such as for cyclists (Poppinga et al. [145]) and for motorbike drivers (Bial et al. [13]).

Pointing has been explored as a metaphor for inputting queries: both Simon et al. [181] and Lei et al. [93] suggest accessing information by physically-pointing the mobile phone in the direction of the corresponding buildings in the real world (see Figure 2.6 (c)). Their validation of this interaction technique has a clear connection to AR, where information queries are also based on pointing the phone's camera in the direction of corresponding objects in the environment. A similar technology is now also produced by GeoVector and commercially available [41].

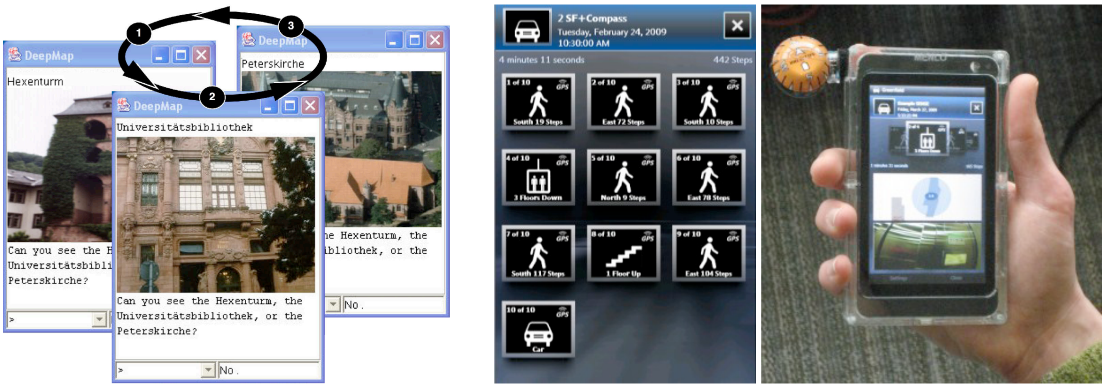
The need to offload users' attentional resources during navigation has a direct impact on our research. Since our focus is on visual AR, we question what are the most important situations where users need to operate the AR interface, and where it is sufficient to support users with less attention-demanding interfaces. In Chapter 7, we present an exploratory evaluation centred on these questions, in an outdoor-navigation scenario.

2.2.3 Exploiting human abilities

Research on handheld navigation systems is starting to look at how human navigation abilities can be exploited to compensate for technological shortcomings. Humans are usually able to navigate the environment by themselves: handheld navigation systems can be therefore designed to support and enhance such human abilities, rather than replace them.

For example, in the context of an indoor navigation system, Butz et al. [19] suggest adapting the amount of information on the screen depending on the localisation accuracy: if the navigation system cannot accurately localise the user, it should increase the amount of information visualised, so that users can exploit the presented information to localise themselves. Kray et al. [87] extend the concept even further, proposing a system that asks users for support whenever it can not localise them. If the system loses localisation, it shows the user photographs of the surrounding landmarks and asks her to identify which landmarks are visible from her position (Figure 2.7 (a)). The system can then triangulate the visibility areas of the different landmarks to localise itself within the environment. Schöning et al. [177] also present a handheld navigation system that exploits users' abilities to improve localisation. In their system, users take a photograph of a public map and later use it as a digital map on their mobile device. The problem of correctly registering the photograph of the map to Global Positioning System (GPS) coordinates is outsourced to the user, who is asked to manually tell the system where the "you-are-here" mark is on the map (the mark is then automatically matched with the phone's GPS position). Löchtefeld et al. [102] present a similar system for indoor maps, working with dead-reckoning rather than GPS.

Brush et al. [18] discuss a handheld navigation system that works completely without localisation technology. Their system is based on *activities* (Figure 2.7 (b)), which are user-centric turn-by-turn instructions (e.g., "walk 30 steps north"). In a user evaluation,


DeepMap

Universitätsbibliothek

Can you see the Hexenturm, the Universitätsbibliothek, or the Peterskirche?

No

2 SF+Compass
Tuesday, February 24, 2009
10:30:00 AM

4 minutes 11 seconds 442 Steps

(a) (b)

Figure 2.7: Exploiting human abilities in handheld navigation systems. (a) Kray et al. [87] ask users for information on landmark visibility, to localise them when GPS does not work. (b) Brush et al. [18] explore the usage of activity-based instructions to support indoor navigation when localisation is absent.

they show that users are able to navigate a path with only a static list of such activities and a magnetic compass. Robinson et al. [161] also investigate how users navigate with only minimal information. In their work, Robinson et al. only provide users with vibrotactile feedback that indicates the direction of the destination and the amount of possible path choices to reach it. In a user evaluation, they show that users are able to reach the final destination with such minimal information. Interestingly, providing only information on the direction of the final destination is exactly what many handheld AR navigation systems do (in particular, AR browsers). In Section 2.3.3, we discuss some related work that looks at how users navigate with an AR system that only shows the direction of the final destination.

Since AR poses high demands on the technological requirements of a navigation system, the possibility of outsourcing some processing to the user when a failure occurs (for example, when tracking breaks) is clearly an interesting fallback. In Chapter 8, we investigate how AR can be complemented by human navigation abilities when tracking does not work. We employ a scenario of indoor wayfinding, where continuous tracking is particularly hard to achieve.

2.3 Augmented Reality Navigation Systems and Their Usability

The work presented in Section 2.2.1 shows the potential of AR for enhancing navigation with first-person views. However, as we also discuss in that section, usability questions arise in the use of AR to enhance navigation systems. Overall, the usability and effectiveness of AR in navigation systems cannot be taken for granted, but it has to be carefully evaluated with proper user studies.

In this section, we present related work in AR navigation systems. As we will see, related work in the field often addresses the problem from a technical perspective, while only a small part of it questions the actual usability of the systems. At this stage, we therefore remind the reader of our three research questions for this thesis, as we will then discuss how the related work addresses these questions:

Q1. When does AR enhance handheld navigation systems?

Q2. When does it fail to enhance them?

Q3. In the latter cases, how can we complement AR with other interfaces to build a hybrid and more effective interface?

The research questions clearly cover a broad research field, and the advantages and disadvantages of AR cannot be properly studied with only a single large user evaluation. As already discussed in Chapter 1, in this thesis we subdivide the research field into four sub-topics: this allows us to explore the impact of AR on smaller subsets of navigational tasks, with sufficiently deep and focused experimental designs. We divide exploration and wayfinding tasks: as we previously discussed in Section 2.1, these two groups cover the space of tasks mainly involved in people's navigation. With respect to exploration, we separately study *egocentric exploration* and *exocentric exploration.* These reflect the two possible ways in which AR can be used to support navigation: the former refers to augmentations placed directly in the environment, whereas the latter refers to augmentations placed on paper maps (or other props). With respect to wayfinding, we separately consider *outdoor* and *indoor wayfinding*: these are the two possible wayfinding scenarios and, despite their similarity, indoor wayfinding is typically more complex in terms of tracking infrastructure necessary for locating the user in the building – we therefore dedicate two distinct chapters to them.

In the following, we adopt the same subdivision for discussing the related work: for each of the four tasks, we first present the interface designs used by other researchers to support the task and then present related usability studies. We conclude each section highlighting the usability questions that have not yet been addressed, as they define the starting point for our research work in this thesis.

2.3.1 Egocentric exploration

In the case of egocentric exploration, the environment becomes an anchor for geo-referenced hypermedia databases and users can explore the information naturally, by physically moving in the environment. A pioneering work in this field is the Touring Machine, first presented by Feiner et al. [32, 33] and later evolved into MARS (Mobile Augmented Reality System) by Höllerer et al. [61–63] (Figure 2.8 (a)). The Touring Machine and MARS allow users to browse a geo-referenced hypermedia database related to the Columbia University campus through AR. Users can navigate the campus and interact with the digital content overlaid on the physical buildings, labeled with virtual information shown through a head-worn display. The authors intentionally label whole buildings and not smaller building features, so that tracker inaccuracies do not affect the usability of the application. The wearable setup of MARS is used in combination with a handheld device, which provides contextual information as a web page. While

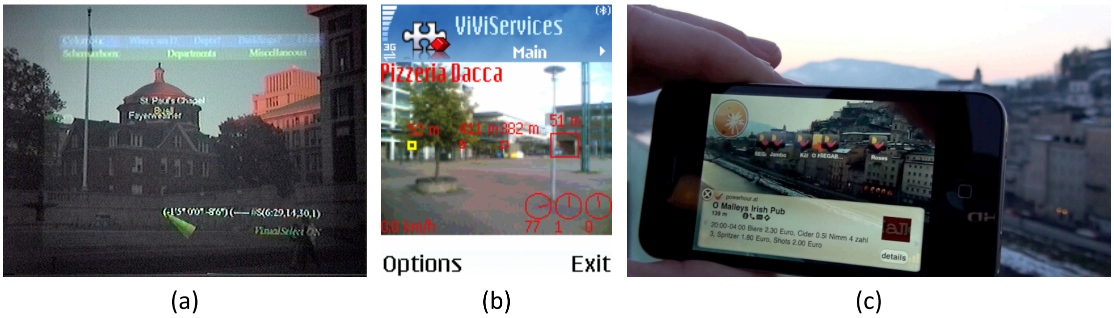
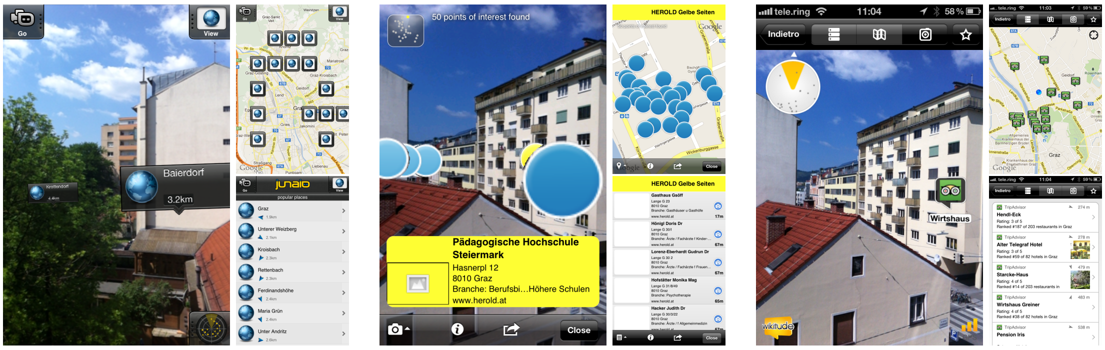


Figure 2.8: Egocentric exploration of information in augmented environments. AR navigation systems typically augment physical objects in the environment with short text labels: (a) MARS by Höllerer et al. [63], (b) MARA by Kähäri et al. [88], (c) Wikitude [207].

the AR context is useful for intuitive exploration and selection of annotations in the environment, it is reasonable to provide details on-demand with a more appropriate interface – in the case of MARS, the interface designers chose to use a web browser. Reitmayr et al. [158] also show a navigation system in which users can explore detailed information regarding the buildings and tourist attractions in their surroundings by simply looking at the corresponding buildings. The usability of all these systems was not formally evaluated in realistic outdoor settings.

The MARA project by Kähäri et al. [88] is the first to implement the concept of the Touring Machine on a mobile phone (Figure 2.8 (b)). MARA provides AR annotations related to the points of interest in the surroundings of a mobile user. Clicking a button on the phone while pointing the camera towards a point of interest shows further information about it. This same concept is recently implemented by a number of commercial AR browsers available on the app stores, for example Wikitude [207] (Figure 2.8 (c)), Layar [91] and Junaio [110]. AR browsers are applications that retrieve geo-referenced content from online databases and present such content to a mobile user on their phone through an AR interface. Most AR browsers (Figure 2.9) augment the environment with annotations, whose distance is typically mapped to the size of the icon. Selecting an annotation typically opens a web browser (or another hypermedia-based interface) that provides further details on the selected annotation. Besides the AR view, AR browsers also provide map and list views that give a better overview of the surrounding information.

The augmentations can present not only static points of interest, but also dynamic content: for example, the Battlefield Augmented Reality System (BARS) by Julier et al. [73] and Livingston et al. [97, 99] focuses on supporting situation awareness for soldiers and informing them about the location of personnel, vehicles and other occluded objects in the



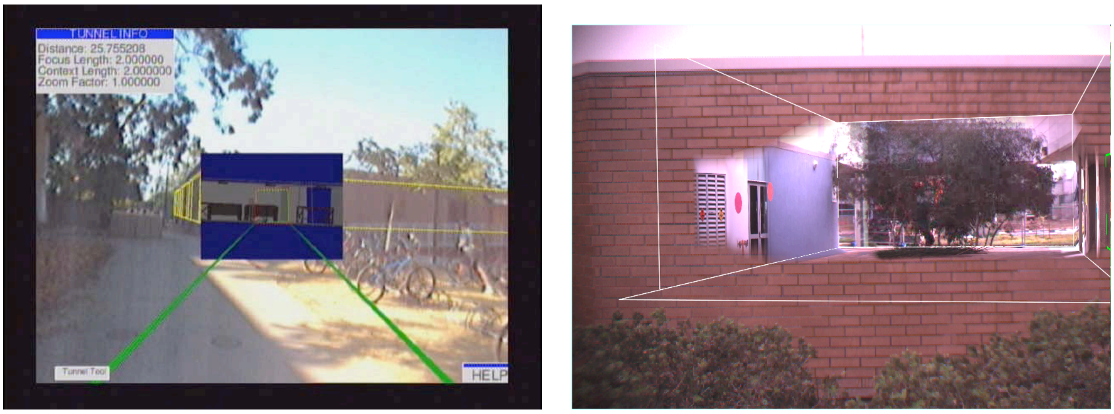
(a) (b) (c)

Figure 2.9: User interface of modern AR browsers. All browsers combine an AR view with map and list views: (a) Junaio [110], (b) Layar [91], and (c) Wikitude [207].

soldiers' view. In BARS, AR is used to support navigation by showing soldiers the position of other moving elements in the environment. The use of AR in this context is justified by three key reasons [73]: first, the urban environment is inherently 3D and not easily represented by a 2D map; second, accessing secondary sources of information requires the soldiers to switch their attention from the environment, which is clearly undesirable; finally, the information to be displayed is often dynamic (e.g., the position of snipers), and can easily be shown by digital technology.

One further advantage of using digital technology is that the AR view can be personalised based on the user's needs and interests. In the context of BARS, Julier et al. [72] also discuss methods for reducing information overload, by filtering out unneeded visual information based on the current mission, the soldier's goals, and physical proximity. We can also consider the results from BARS outside the military context, highlighting how AR can be used to show the dynamic position of moving objects, for example friends: Kähäri et al. [88], in their MARA project, use AR to show the position of other MARA users in the surroundings.

Since annotations are merged with live images from the video camera, AR is bound to the frame of reference of the video camera. In the case of handheld navigation systems, the camera is physically bound to the egocentric street-level view of the user. As we discussed in Section 2.2.1, users often prefer higher viewpoints (for example, bird's eye views), because they provide a broader overview of what is in their surroundings. In contrast, with egocentric exploration in AR the amount of information visible from the viewpoint of the camera can be insufficient, either because of occlusions between the camera and the information, or because the information is outside the current camera view. As shown in

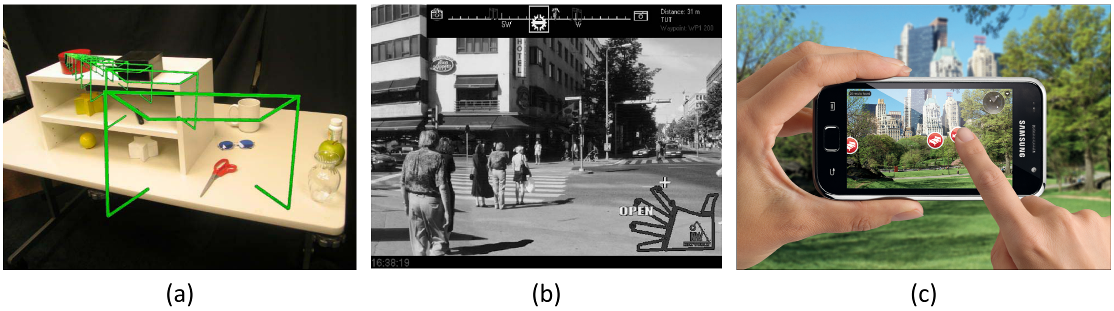


(a) (b)

Figure 2.10: Egocentric exploration of occluded information. See-through interfaces for AR navigation systems: (a) tunnel metaphor by Bane et al. [9], (b) x-ray vision by Avery et al. [7].

a user study by Fröhlich et al. [39], a usability issue is that users need to physically turn around a lot, in order to find the information. This highlights the value of an interface that combines AR with third-person views such as 2D maps, 3D maps or VR.

Supporting users in the case of occluded augmentations has been already extensively explored. Various navigation systems employ transparency and x-ray vision to communicate depth and occlusion of annotations. Livingston et al. [98] conducted an experiment to evaluate various transparency cues to communicate multiple levels of occlusion. Their results show that a ground plane is the most powerful cue, but rendering occluded objects in wireframe and filled with a semi-transparent colour is also a good cue. In a later experiment, Tsuda et al. [193] obtain analogous results. Bane et al. [9] discuss a technique for x-ray vision in a mobile context, but the authors do not conduct a user evaluation of the technique. A tunnel metaphor is used to browse the rooms of a building from the outside, and a wireframe rendering of the tunnel provides cues about the depth of the various rooms (Figure 2.10 (a)). Avery et al. [6, 7, 139] show a similar x-ray vision technique that employs transparency and cut-outs to communicate multiple layers of occlusion (Figure 2.10 (b)). Schall et al. [172] and Zollmann et al. [216] apply x-ray techniques in the context of exploration of underground infrastructure. Wither et al. [209] conduct a study of various depth cues for exploring information in AR, showing that adding a non-AR top-down view enhances depth understanding. Sandor et al. [169, 170] recently suggest a metaphor different than x-ray vision: they use a 3D model of the city to virtually melt the closest buildings, to show the occluded content behind them. However, the authors do not conduct a user evaluation of the technique. In summary, the evaluations suggest



(a) (b) (c)

Figure 2.11: Guiding users to the information outside the current view. In the related work, different interfaces are applied for this purpose: (a) the attention funnel by Biocca et al. [14], (b) the context compass by Lehikoinen et al. [92], (c) a radar-like interface in Layar [91].

that rendering a ground plane, various transparency and see-through effects, as well as providing an alternative top-down view are effective depth cues in AR.

Supporting users in the case of augmentations outside the current camera view has not been studied so extensively, and is the topic of our work presented in Chapter 5. Biocca et al. [14] present the attention funnel (Figure 2.11 (a)), an AR visualisation element shaped as a tunnel which guides the attention of a user towards a specific object in the environment. The authors evaluate the technique in a head-worn setup, comparing it against visual highlighting (a 3D bounding box) and a verbal description of the object. Results show that the attention funnel reduces visual search time and mental workload. However, the interface also provides visual clutter, so the user should be able to disable the tunnel when it is not needed. Schwerdtfeger et al. [175, 176] apply the technique to the context of order picking in a warehouse, but their evaluations are limited to different designs of the interface and not comparative against other navigation interfaces (for example, a map). In general, all evaluations of the attention funnel are conducted in small indoor settings, and there is no clear indication of the effectiveness of such interface outdoor, or in comparison to more established interfaces such as maps. Furthermore, it is not clear if the tunnel metaphor, which works well in head-worn setups, can also work well on handheld devices.

AR is also often combined with other non-AR interfaces that support users in understanding where off-screen augmentations are located. For example, Lehikoinen et al. [92] propose a 2D overlay interface called the Context Compass (Figure 2.11 (b)), very similar to the 2D overlay used in Tinmith by Thomas et al. [189]. This interface uses a compass metaphor to show the horizontal orientation of annotations with respect to the user. It is a linear and user-centred indicator of orientation: icons in the centre of the overlay represent annotations currently visible by the user, whereas icons to the side of the

Problem	Solution	User evaluations		
Occluded augmentation	X-ray view / transparency	Livingston et al. [98], Tsuda et al. [193]		
Off-screen augmentation	3D AR cue	Biocca et al. [14], Schwerdtfeger et al. [175, 176]	Schinke et al. [174]	**Our contribution (Chapter 5)**
	2D overlay	Lehikoinen et al. [92]		
	Transition to another interface	Güven et al. [47, 48]		

Table 2.1: User evaluations in egocentric exploration, and placement of our contribution. Previous work focuses mostly on one specific solution for off-screen augmentations. In contrast, we sistematically study different solutions for providing overview of off-screen augmentations, for various types of user tasks.

overlay represent annotations outside the field of view of the user. The Context Compass is only evaluated in a small-scale indoor setup, and not in comparison to other interfaces. Güven et al. [47, 48] discuss moving between AR and VR modes for browsing distant and occluded hypermedia objects. A user evaluation shows that the VR techniques are slower than a transparency-based interface, but more accurate. Höllerer et al. [62] also use a design that transitions between AR and VR, to allow users browsing remote camera views on demand, but they do not conduct an evaluation of their technique. Sandor et al. [169] uses a 3D virtual model of the environment to provide a distorted view of the surroundings, with a much larger field of view than the used camera. This technique is partially related to our Zooming Panorama technique (presented in Section 5.2) but the authors do not conduct any usability evaluation. Most AR navigation systems use a map or a World-in-Miniature (WIM) (a 3D miniaturised version of the surrounding environment, first presented by Stoakley et al. [187]) to provide overview, either on a separate handheld device that can be brought into view whenever needed (as done, for example, by Höllerer et al. [63] and by Reitmayr et al. [157]) or by tilting the view down (as down, for example, by Bell et al. [12] and by Kähäri et al. [88]). The tilting motion is an established way to space-multiplex AR and other interfaces: users look up for the AR view and look at the floor, where there is usually no augmentation, to access another interface. We will also use tilting motions for the same purpose in our work, in Section 5.2, 7.1 and, in a slightly different form, in Section 8.2 and 8.3. In commercial AR browsers, radar-shaped overlays and 2D maps (Figure 2.9, Figure 2.11 (c)) are also often used to show the position of all the surrounding augmentations. In parallel with our research work, Schinke et al. [174] conducted a first comparative evaluation of 3D AR arrows against a radar-shaped overlay interface. Their evaluation shows that AR arrows outperform the radar overlay in accuracy, when users are asked to estimate the physical direction of off-screen annotations.

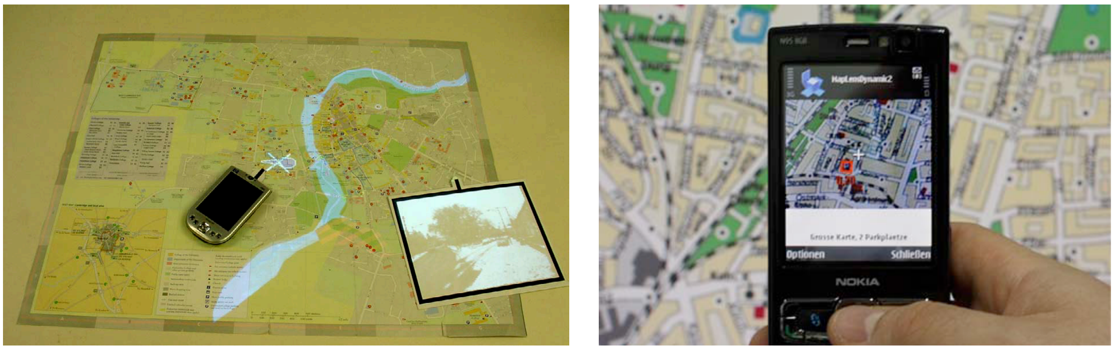
In general, however, there is no systematic evaluation of the usability of AR and non-AR cues for off-screen augmentations for different user tasks (see Table 2.1). In Chapter 5, we present a broad research on the usability of AR for egocentric exploratory tasks, studying not only how AR supports search tasks where one single augmentation is involved (Section 5.1) but also how to support AR with other interfaces when users need an overview on multiple surrounding augmentations (Section 5.2).

2.3.2 Exocentric exploration

Exploring detailed digital information by simply pointing a mobile device to different locations on a paper map has already been studied outside the AR community, for example in the pioneering work by Fitzmaurice et al. [35], or more recently by Norrie et al. [131] and Reilly et al. [153]. The last two works show evaluations both in controlled and real-world settings: their results highlight the value of the mixed-media approach of augmenting physical maps.

Within the AR community, the feasibility of augmenting paper maps has been largely validated for small-size maps, for example by Bobrich et al. [15] and Hagbi et al. [49], as well as for larger table-size maps, for example by Ishii et al. [68], Reitmayr et al. [156] (Figure 2.12 (a)) and Olwal et al. [132]. Moore et al. [115] exploit the tangible nature of paper maps, using a 3D paper cube as an interface for exploring a potentially infinite augmented map. Martedi et al. [104] further consider the tangible nature of paper maps, proposing tracking and interaction solutions for partially folded maps.

The first related work to study how users operate a handheld navigation system based on augmented maps is the one by Schöning et al. [178] and Rohs et al. [165, 166] (Figure 2.12 (b)). The focus of these studies is on how users explore the map, and the systems are not validated outdoors in real-world navigation tasks. The results of the evaluations are very encouraging for the use of augmented maps, showing that the key advantage of using an augmented map over a digital map lays in the physical interaction with the map [166], which allows for faster and more efficient exploration of the map. The visual context provided by the physical map behind the mobile device has a smaller impact than the proprioceptive feedback give by the physical movement of the user when hovering the device over the map; this is also due to the fact that switching attention between device and paper map is cumbersome. Furthermore, having a paper map becomes more advantageous when the map is large and the information sparse on it [165]. Schöning et al. [179] present a further iteration of their augmented-map interface that uses a pico-projector, rather than



(a) (b)

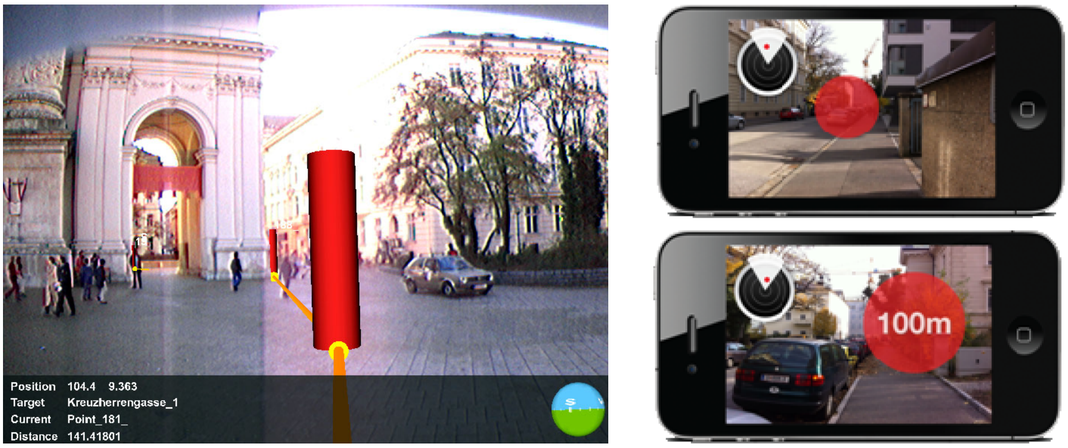
Figure 2.12: Exocentrix exploration with augmented maps. (a) Reitmayr et al. [156] investigate tabletop augmented maps and tangible interaction with the map for details on demand. (b) Rohs et al. [165, 166] conduct deep quantitative studies on the usage of augmented maps for exploring information.

a magic-lens metaphor. An evaluation of this new design shows that projector-based AR speeds up search tasks compared to a magic lens, but the projector-based solution is not studied under outdoor lighting conditions. Overall, all these works study in detail how users operate augmented maps in controlled settings, but lack insight on how users would really use an augmented map in real-world settings and for outdoor navigation tasks.

The first to study augmented maps in outdoor settings are Morrison et al. [118]. Their experiments confirm the benefit of using an augmented map over a digital map for easily browsing information, but also show the value of a paper map as a physical prop that fosters collaboration and communication between multiple users. Two limitations of this work are the weak tracking technology, which forces users to operate the augmented map solely after laying the map onto a stable surface, and the fact that users were forced to share one single mobile device – in real-world settings, we can assume that multiple users will typically all have their own phone. In Chapter 6, we extend the work by Morrison et al., in collaboration with the original authors. In our work, we study how stable tracking and multiple devices change the way users operate augmented maps for exploring information in real-world outdoor settings.

2.3.3 Outdoor wayfinding

Besides the use of AR for exploring information, users can also receive support for wayfinding to a specific destination, by visualising the path from one location to another in AR. A pioneering work in this field is Tinmith by Thomas et al. [189], later refined by Piekarski et al. [137, 138]. In Tinmith, a user explicitly defines a desired path as a finite sequence



(a) (b)

Figure 2.13: Outdoor wayfinding in AR navigation systems. (a) Reitmayr et al. [158] guide users by augmenting the environment with a set of viewpoints. (b) Rehrl et al. [152] show the final destination and its distance from the user as an augmentation in the environment.

of waypoints: Tinmith then interactively shows users the position of the next waypoint as they navigate the environment through a head-worn display. Tinmith was only evaluated in the context of collaborative navigation [185, 186], but the focus of the evaluations was on different cues for collaboration rather than on wayfinding performance. MARS [63] also supports a collaborative wayfinding mode between an outdoor user and a remote guide: the remote guide can sketch a path to appear in the AR view of the outdoor user, to support her wayfinding. In a recent work, Reitmayr et al. [158] show an interesting AR navigation system for collaborative wayfinding: for example, the system can be instructed to guide multiple users so that they can meet halfway between their start locations. All waypoints are visualised in the environment and connected to each other by arrows (Figure 2.13 (a)).

The value of AR as a support for wayfinding is also largely shown in the automotive field, typically considered a more attentional-demanding – and thus more critical – field than pedestrian navigation. Numerous AR cues have been designed and evaluated in driving simulators. Kim et al. [78] and Medenica et al. [107] show that AR improves attention on the driving task compared to traditional in-car navigation systems, but in both cases the comparison is conducted between a heads-up AR condition and a heads-down non-AR condition, thus the effect of AR vs. non-AR is hardly separated from the effect heads-up vs. heads-down. Tönnis et al. [194–196] show that in-car AR cues improve drivers' performance, but Plavšić et al. [144] highlight a usability issue when augmentations are outside the current view of the driver, in which case a non-AR overview is preferred. Fröhlich et al. [37] are the first to perform an evaluation of in-car AR wayfinding cues outside a simulated environment, with users driving on the motorway. They compare audio instructions, a map interface, and an AR interface similar to the one presented by Tönnis et al. [196]. The results are more conservative than those from simulator-based studies, suggesting that AR does not generally improve drivers' performance, but it is beneficial for communicating instructions in emergency settings (for example, for telling the driver to stop immediately, and communicating the lane in which she has to stop).

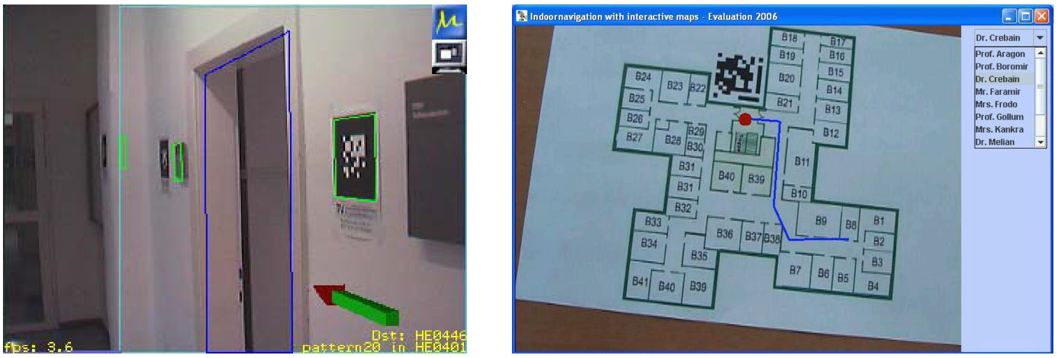
AR cues have also been validated in aviation, a more demanding scenario in which reducing the pilots' workload is uttermost important. In aviation, it is common to use *pathway displays*, visualisations which augment the cockpit view with virtual tunnels that the pilot has to follow. A good overview of pathway displays is given by Newman et al. [129]. It is now generally acknowledged that tunnel-like AR cues can improve flight performance, as shown for example by Kramer et al. [85]. Foyle et al. [36] also apply a similar concept to surface aircraft taxi operations, but do not conduct an evaluation of their system. Similar AR cues are also adopted by commercial handheld navigation systems, such as Augmented Driving [67] and Virtual Cable [184]. However, Wickens et al. [206] highlight the potential risk of cognitive tunneling when using the tunnel cues in an aircraft, with the risk that the pilot might miss important information located outside the AR display. Tönnis et al. [195] discuss the same issue for AR cues in cars. It is unclear if this issue applies also to handheld devices, which are not always-on screens like the heads-up displays in cars and airplane cockpits. However, this issue poses safety concerns that should be taken into consideration when designing AR navigation systems.

In the context of handheld navigation systems for pedestrians, the same AR solution used for car navigation system and head-worn displays is often applied (for example, in the work by Narzt et al. [127] or in Wikitude Drive [208]). It is however arguable that pedestrians are willing to constantly use AR to navigate (and constantly hold the phone upwards): this is already a theme of discussion in the AR community (e.g., see Tokusho et al. [191]). Indeed, as discussed in Section 2.2.2, a current trend in handheld navigation systems is rather to offload the burden from the visual sense onto other senses, relying on vision only when necessary. In Chapter 7, we present the first real-world evaluation that studies how often, where and how users need to access AR cues for turn-by-turn instructions during outdoor wayfinding. In parallel with our research, the topic was also investigated by Dünser et al. [31] and Rehrl et al. [152]. They both use AR cues different than ours, showing the direction of the users' ultimate destination (Figure 2.13 (b)) rather than turn-by-turn arrows – a cue similar to the vibrotactile one used by Robinson et al. [161] and presented in Section 2.2.3. This AR cue shows increased workload compared to a map-based interface [161], but works better than a map when the destination is within visible line of the user [31]. This result is in line with our findings, presented in Chapter 7, where we found that maps can be sufficient for outdoor wayfinding, while AR cues are most needed for decision within visibility range (for example, the turns to take or the position of the final destination to reach). Our results also differentiate from the related work because our study is exploratory and not comparative, thus leaving the users full freedom in using either the map or AR whenever needed.

2.3.4 Indoor wayfinding

Indoor wayfinding is a more complex scenario than outdoor wayfinding, because there is no standardised interface (as compared to the 2D maps outdoors) or reliable localisation technology (as compared to GPS outdoors). Many researchers look at supporting indoor navigation with AR cues from the technical perspective of how to continuously localise a user in a building. For example, Kalkusch et al. [74] and Reitmayr et al. [157] present an indoor navigation system that uses an AR arrow to show the directions to the next waypoint. Their system is based on black-and-white artificial markers, densely placed on the corridors' walls. Wagner et al. [203] later show how the same indoor navigation system can also run on handheld devices (Figure 2.14 (a)). Piekarski et al. [135, 136] use a similar marker-based approach, while other researchers use more advanced computer-vision tracking (for example, Kim et al. [77] and Hile et al. [55]), or sensor-based tracking (e.g., Tenmoku et al. [188]) and dead-reckoning (e.g., Merico et al. [109]). This work focuses on the technical feasibility of indoor navigation systems, but there is no formal evaluation of the usability of such systems.

Höllerer et al. [64] are the first to discuss the issue of tracking degradation from a usability perspective. Similarly to the work by Butz et al. [19] (discussed in Section 2.2.3), Höllerer et al. suggest transitioning between AR and a WIM view depending on tracking accuracy. When the tracking is sufficiently accurate, annotations and route arrows are superimposed on the environment using AR. When the tracking accuracy degrades, the interface smoothly transitions to a WIM view, and an avatar representation is used to indicate the current position of the user in the WIM. Rather than inaccurately placing the augmentations – which could potentially confuse users – the authors choose to transition the system to a WIM interface that is more robust to tracking inaccuracies,



(a) (b)

Figure 2.14: Indoor wayfinding in AR navigation systems. (a) Wagner et al. [203] continuously guide the users with AR arrows in the environment. (b) Müller et al. [126] support users with augmented maps at sparse locations in the building.

and more informative for the user to navigate. This idea is developed further by Hallaway et al. [50], but the authors do not conduct any user evaluation of the system. This work is tightly related to our work presented in Chapter 8, where we further explore the issue of tracking degradation evaluating the case of complete loss of localisation.

Also similar to our work from Chapter 8, Müller et al. [126] support indoor navigation with augmented public maps at sparse locations in the building (Figure 2.14 (b)). The results from a preliminary evaluation show that the availability of augmented maps improves navigational performance, significantly reducing the amount of navigational errors. These results are consistent with our results. However, Müller et al. do not provide any navigational cue to the user when a map of the building is not visible. In Chapter 8 we extend the concept of augmented information points at sparse locations, conducting a number of evaluations aimed at discovering what type of information users need when departing from such accurate information points and entering areas in which localisation is completely unavailable.

Finally, there are strong commercial efforts for supporting indoor navigation, such as the most recent version of Google Maps [44], which seamlessly merges outdoor and indoor maps, or recent research work on Ovi Maps[*]. However, there is currently no AR navigation system that supports continuous navigation guidance indoors, and the systems that have been demonstrated at large-scale indoor venues are typically based on localisation at sparse locations, such as our system Signpost, presented in Section 8.1, and Junaio [110], which has been demonstrated at the Intel Developer Forum in 2010[†].

[*] Nokia Indoor Navigation: `http://research.nokia.com/news/11809`.

[†] Indoor AR at the Intel Developer Forum: `http://www.youtube.com/watch?v=k3bFBn_Bs8Y`.

2.4 Tracking requirements for AR navigation systems

Accurate tracking is a fundamental requirement for all AR applications, because augmentation is only possible when the position and orientation of the device in the environment can be estimated. Given the importance of the topic of tracking for any AR navigation system, we give here a brief overview of the state of the art in the field. Since the focus of this thesis is on handheld platforms, we put particular emphasis on the challenges involved in performing tracking on handheld devices.

Due to the interactive nature of AR applications, tracking estimates have to be calculated in real-time, ideally at the speed of the camera updates (usually 30 frames per second). Furthermore, the estimates have to be extremely accurate. Azuma et al. [8] argue that tracking accuracy should be a fraction of a degree in orientation and a few millimetres in positioning, in order to achieve reasonable augmentations. While Feiner et al. [33] show that lower accuracy still allows coarsely labelling large buildings, it is typically desirable to position augmentations at a finer grain than that, for example to highlight a building's entrance. Therefore, trackers should typically offer centimetre or millimetre accuracy in position and sub-degree accuracy in orientation, to allow for correctly registered augmentations in navigation systems.

Early navigation systems targeting large-scale outdoor scenarios, such as the Touring Machine [33] or Tinmith [189], perform tracking using sensors. Typically, GPS is used to estimate the position of the user, while an inertial unit and a magnetometer are used for estimating the orientation. Another approach is to incrementally measure users' movements (*dead reckoning*), as shown for example by Hallaway et al. [50] and Löchtefeld et al. [102], but these solutions typically drift over time. Sensor-based solutions are usually inaccurate and unreliable – for example, Feiner et al. [33] report that they can place annotation only on whole buildings, and not on finer details, due to the tracking inaccuracy. While advancements in technology have improved the accuracy of high-end sensors, the sensors available on handheld devices do not provide an accuracy suitable for AR applications. The issue does not only depend on the use of cheap sensors, but also on the need of mounting all parts close to each other, due to intrinsic space constraints, with the resulting problem of interferences. In general, positioning accuracy with assisted GPS on mobile phones is in the range of several meters [212], while orientation accuracy is typically a few degrees [173]. A phone's magnetometer can further deviate due to external factors, for example by simply wearing a metal wristwatch while holding the device [173]. This level of accuracy is clearly a few orders of magnitude worse than the one needed for correct augmentations. Nevertheless, many commercial AR navigation systems on mobile phones (for example, Wikitude [207] or Layar [91]) employ mostly sensor-based tracking and typically produce inaccurate and unstable augmentations.

Vision-based approaches offer promising results for accurate real-time tracking in AR applications. Early work in vision-based tracking uses fiducial markers, artificial tracking targets designed for good tracking performance. Kato et al. [75], for example, developed ARToolkit, a tracking system that works with tracking targets that have a prominent black-on-white border. Möhring et al. [114] present a tracking solution for mobile phones that uses colour-coded 3D markers. Around the same time Rohs et al. [162] create the VisualCodes system for phones. Both works provide only simple tracking of 2D position on the screen, 1D rotation and a very coarse distance measure. Wagner et al. [203] and Henrysson et al. [53] are the first to show the feasibility of accurate 6 degrees of freedom (DOF) marker-based tracking on handheld devices. The accuracy of modern marker-based trackers is typically of a few millimetres and a fraction of degree [134], and therefore suitable for AR.

Instrumenting the environment with artificial tracking targets is impractical in many scenarios, and more recent work looks at natural features as a support for tracking. Point-based approaches use corner detectors and matching schemes to obtain correspondences between 2D locations in the video image and known 3D locations in the environment. For example, Skrypnyk et al. [182] describe a classic system based on the SIFT descriptor [100]. Lepetit et al. [94] recast matching as a classification problem, using a decision tree. This work was later improved by Ozuysal et al. [217], while further reducing the necessary computational work. Despite the efforts on reducing the computational load of natural-feature tracking, such approaches were considered too demanding for handheld platforms until recently. Our work [201, 202], presented in Chapter 3, is the first one to show the applicability of point-based natural-feature tracking of planar targets on handheld devices. Very recent commercial products for handheld devices, such as Vuforia [149] or Junaio [110], incorporate natural-feature tracking solutions for planar and 3D objects, ultimately bringing this enabling technology outside the research labs. The accuracy of natural-feature trackers is typically presented in pixels, and state-of-the art trackers can achieve an accuracy of one pixel or better [202] – the metric value of such accuracy depends on the distance of the camera from the tracking target. Independent of the metric accuracy, the ability to position augmentations with pixel accuracy clearly makes this tracking technology suitable for AR.

In larger-scale scenarios, such as outdoor navigation, the need for accuracy brought research in the direction of hybrid tracking solutions, which combine sensor measurements with computer-vision techniques. For example, Reitmayr et al. [154, 155] show a system that benefits from combining vision-based edge tracking with sensors. Their system uses GPS to obtain an initial estimate of the user's position in the environment; edge tracking is then used at runtime for accurate augmentations, while an inertial unit helps coping with fast camera movements or occlusions that occur during normal usage. Klein et al. [79] also combine accurate edge tracking with gyroscope measurements, for higher robustness. These systems use a model-based tracking approach and therefore require previous knowledge on the appearance of the outdoor environment, which is not always possible.

Indoor scenarios are even more challenging, because GPS is typically unreliable due to bad signal reception, while magnetometers often suffer from the strong magnetic influences. Different types of sensing infrastructure have been used as a basis for indoor localisation in AR, for example wireless or infrared tracking [50]. Position accuracy typically varies between about 1 metre for infrared tracking [50] and several meters for wireless tracking [105], while orientation must be usually estimated using sensors. High-end ultrasonic or infrared trackers offer an accuracy of a few millimetres, but they can cover only a small volume in the size of a few metres and are therefore not suitable for indoor navigation. Wagner and Schmalstieg [203] use computer vision to localise users, by detecting markers mounted on the walls. Overall, all these solutions require instrumentation of the environment and knowledge about the location of emitters/receivers or of artificial markers in the environment, which is not always a viable solution in large-scale environments.

An interesting hybrid tracking approach is taken by adaptive tracking systems, which dynamically select the best tracking technology available depending on the user's location, as done for example in the Ubitrack project by Pustka et al. [148] or by Hallaway et al. [50]. An analogous approach is also used in commercial map-based products: for example, both Apple[‡] and Google[§] combine GPS, cell ID and Wi-Fi positioning in their map applications. In this case, the accuracy of the tracker varies over time, depending on the technology available.

If no model of the environment is available, sensor-based trackers can still benefit from

[‡] Apple Q&A on Location Data: `http://www.apple.com/pr/library/2011/04/27Apple-Q-A-on-Location-Data.html`.

[§] Google, Location source and accuracy: `http://support.google.com/gmm/bin/answer.py?answer=81873`.

the integration of frame-by-frame computer vision methods. For example, You et al. [211] combine sensor estimates with optical tracking to stabilise the augmentations. Jiang et al. [69] discuss using gyroscopes for a first orientation estimate and frame-by-frame edge tracking for refining such estimate. Similarly, Satoh et al. [171] and Ribo et al. [159] also combine sensor tracking with point-based natural-feature tracking. The results of Satoh et al. show strong improvements in registration accuracy even after 40 minutes of usage, highlighting the value of vision-based refinements of the sensor-based estimates. Wagner et al. [199] discuss the use of frame-by-frame optical flow on handheld devices, for improving tracking robustness. In a recent work [200], presented in Chapter 4, we show a vision-based system that constrains the tracking problem to 3 DOF (only orientation), thus achieving very robust real-time execution on a mobile phone. In a further work [173], also discussed in Chapter 4, we show how this system can be combined with sensors, to obtain very robust hybrid orientation tracking in a world-aligned reference frame.

More recent research on the topic of tracking in unknown environments goes beyond frame-by-frame vision tracking and aims at building a model of the environment on the fly. Knowledge is borrowed from the field of robotics, where Simultaneous Localisation and Mapping (SLAM) approaches were first proposed [183]. The key idea behind SLAM is to extract natural features from the environment and store them in a 3D map, which is incrementally refined over time as the user moves in the environment. SLAM-like systems have been applied to AR by various researchers, for example Klein et al. [80] and Davison et al. [29] with a monocular camera, or by Newcombe et al. [128] using a depth camera. In recent works, Klein et al. [81] and Pirchheim et al. [143] demonstrate the feasibility of SLAM also on handheld devices. The research on SLAM clearly opens up novel possibilities for AR applications, as any unknown environment can potentially be augmented.

However, in contrast to model-based tracking, a semantic link to the tracked scene is typically missing in SLAM systems, because the system has no knowledge about the meaning of the features that are being tracked. Furthermore, SLAM-based systems only track the device in an arbitrary initial reference frame, which is not aligned to world coordinates. A separate and complementary branch of research looks exactly at the problem of localising from an image (for example, Li et al. [95], Chen et al. [21], or the recent project Read/Write World [112]). These methods are an important and very promising complement to tracking in unknown environments, as they can serve to provide an initial estimate of the device's position and orientation in the environment, which the device can then exploit for aligning the SLAM coordinate system to world coordinates. In the context of AR, Arth et al. [4] show a system that runs on mobile phones and can localise users with sub-meter accuracy in a wide-area scenario, both indoors and outdoors. Their system exploits an offline-generated point cloud, divided into GPS-referenced feature blocks, to perform on-device localisation. In a more recent work, Arth et al. [3] also show how this localisation system can be combined with our orientation tracker [200] (Chapter 4), successfully combining accurate localisation in world-coordinates with real-time tracking of the device. Overall, combining a model-based localisation system with an accurate tracker that does not need a model of the environment is a very recent development in the field. It will certainly be a key technology for handheld AR navigation system in the upcoming years, once it will be proven to be sufficiently robust for end-user applications.

2.5 Summary

In summary, while there are several examples of AR navigation systems, there is significantly less work on the usability of AR in such systems. In particular, there is little evidence that AR is an effective enhancement to navigation systems. Furthermore, most user evaluations are conducted in laboratory studies, while there is a lack of validations on real-world navigation tasks. In this thesis, we combine laboratory studies with evaluations in real-world settings, to gain novel insight into the usability of AR navigation systems in realistic scenarios. This is the core contribution of this thesis and is presented in Chapter 5–8. In order to evaluate our work in real-world settings, we faced the need of robust and accurate tracking: as a secondary contribution of this thesis, we implemented two novel trackers. These are fundamental for the robust functioning of our interface prototypes. We briefly present these two trackers in the following part of this thesis, in Chapter 3–4.

PART II

Tracking technology

Chapter 3

Tracking planar targets

In this part of the thesis, we briefly discuss our contribution to the field of mobile AR tracking. A new tracking solution represents a fundamental enabling technology for our interface-design work on AR in handheld navigation systems, as it typically opens up new possibilities for interaction with the augmentations. Furthermore, robust tracking is necessary to develop prototypes that must be operated by end users and work robustly in real-world conditions. During the course of this thesis, the author contributed to a number of tracking systems, which we present in this chapter and in the following one. Given the focus of this thesis, we try to give a brief overview of the technical contribution of each tracking system, while putting more emphasis on how the system can be used to build new interactions for handheld navigation systems.

In this chapter, we present our work on natural-feature tracking of planar targets. This relies on real-world elements, rather than on artificial targets, to accurately calculate the position and orientation (*pose*) of the user's device in the environment. This gives handheld navigation systems the possibility of augmenting everyday's life planar objects, such as paper maps, with accurately registered AR cues.

Natural-feature tracking is computationally demanding and only recently became feasible on handheld platforms. Our natural-feature tracking system is the first system capable of tracking planar targets with 6 DOF at real-time frame rates (30Hz) on modern handheld platforms, using only the computational resources and the built-in camera of the handheld device. In the following, we give an overview of the functioning of our tracking system and present evaluation results on its robustness and performance. For further details, we refer the reader to our publications [201, 202].

3.1 Tracking system

Our tracking system relies on two separate trackers: *PhonySIFT* and *PatchTracker* (Figure 3.1). PhonySIFT does *tracking by detection*: for every video image, it detects all the natural features in the image, matches them against a database of known features and finally estimates the camera pose. Frame-to-frame coherence is not considered in the tracking process. In contrast, PatchTracker does only *active search*: based on a motion model, it estimates where to look for known features and how they will appear due to locally affine transformations.

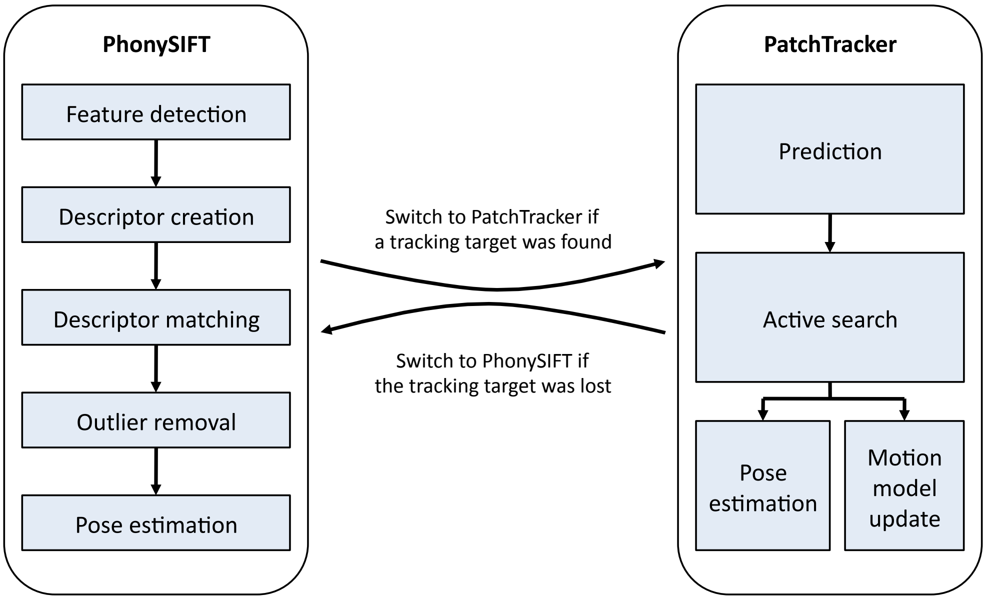


Figure 3.1: Runtime pipeline of our natural-feature tracking system. We combine two separate trackers, PhonySIFT and PatchTracker.

PatchTracker requires another tracker to start, because it cannot initialise without a previously known pose. In our tracking system PhonySIFT is used for initialisation. As soon as PhonySIFT detects a target and estimates a valid pose, it hands tracking over to PatchTracker. PatchTracker uses the pose estimated by PhonySIFT as a starting pose, and then uses its own poses from frame to frame for continuous tracking. In typical application scenarios, PatchTracker works for hundreds or thousands of frames before it loses the target. In such case, our tracking system switches again to PhonySIFT, for reinitialisation.

3.1.1 PhonySIFT

PhonySIFT derives its name from a tracking approach called Scale Invariant Feature Transform (SIFT), originally proposed by Lowe [100]. PhonySIFT is a model-based tracker, like the original SIFT, and therefore requires previous knowledge of the tracking target. In the next sections, we present both the runtime tracking pipeline and the offline step, which is necessary to create a database of features for a given tracking target.

3.1.1.1 Runtime pipeline

As shown in Figure 3.1, the realtime pipeline of PhonySIFT combines a sequence of steps: feature detection, descriptor creation, descriptor matching, outlier removal, and finally pose estimation. The next sections detail each of these steps separately.

Feature detection. We detect features using the FAST corner detector from Rosten et al. [167] with non-maximum suppression. FAST is known to be one of the fastest corner detectors, while still providing good repeatability [40]. Since FAST does not estimate a feature's scale, we store feature descriptors from multiple scales of the video frame. We build an image pyramid, each level scaled down with a factor of $1/\sqrt{2}$ from the previous one, and calculate FAST corners at each scale level (Figure 3.2).

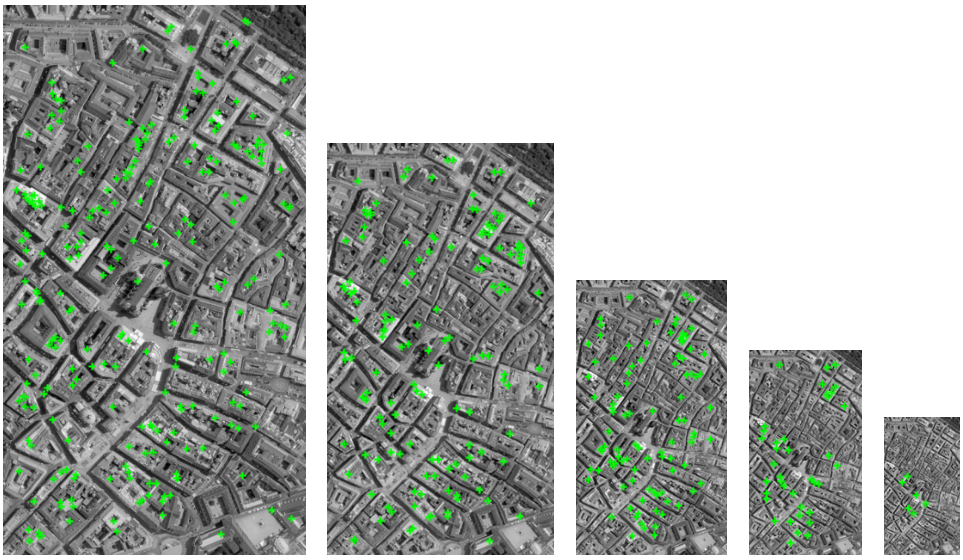


Figure 3.2: Feature detection in PhonySIFT. Since FAST [167] does not estimate the features' scale, we extract features at multiple scale levels.

Descriptor creation. We first blur the patch around a feature with a 3×3 Gaussian kernel. Like in the original implementation, we then estimate feature orientations by calculating gradient direction and magnitude for all pixels of the kernel, and use them to assign one or more orientations to the feature. For each feature and each orientation, we then rotate the image patch surrounding the feature with sub-pixel accuracy (Figure 3.3). We finally create a 36-element SIFT descriptor using 3×3 sub-regions with 4 gradient bins each. As Lowe outlines in his paper [100], this combination performs only ~10% worse than the best variant.

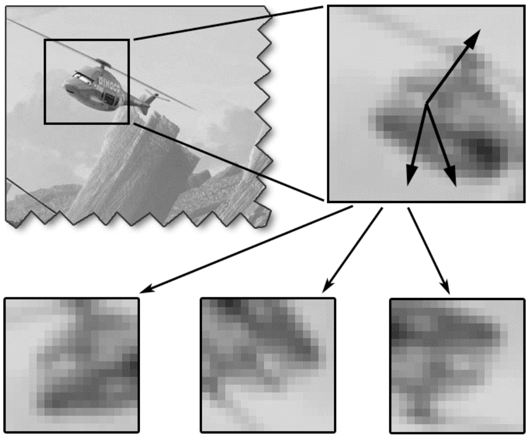
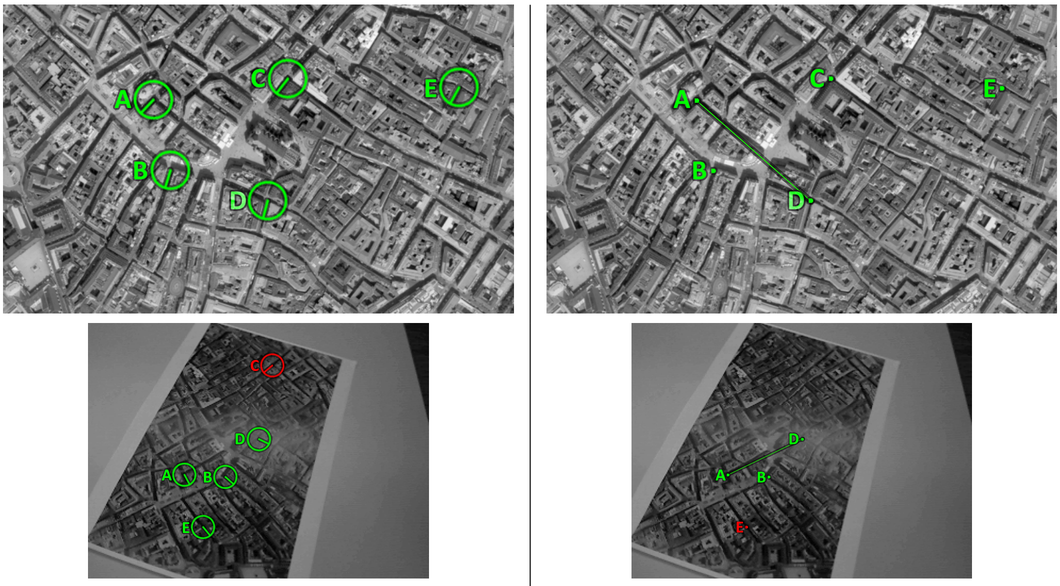


Figure 3.3: Patch rotation in PhonySIFT. For the three main orientations of the feature, we rotate the surrounding image patch with sub-pixel accuracy before calculating a feature descriptor.

Descriptor matching. After creating the descriptors for all features in the camera image, we match them against a database of known descriptors for a specific tracking target (see Section 3.1.1.2). To speed up the matching, we use *spill trees* [96], a variant of k-d trees that uses an overlapping splitting area. We combine a number of spill trees into a *spill forest*: multiple spill trees with randomised dimensions for pivoting allow for a highly robust voting process, similarly to the randomised trees [94]. Since only a few values of a descriptor are expected to be wrong, a descriptor has a high probability of showing up in the "best" leaf of most trees in the forest. We therefore visit a single leaf in each tree and merge the resulting candidates. Descriptors that show up in only one leaf are discarded. All others are matched using Sum of Squared Difference (SSD).

Outlier removal. Although SIFT is known to be a very strong descriptor, it still produces outliers that have to be removed before the pose estimation step. Our outlier removal works in three steps. The first step uses the feature orientations: we correct all relative feature



Orientation test

Geometric test

Figure 3.4: Outlier removal tests in PhonySIFT. Before running a RANSAC-based outlier removal, we conduct two simple and fast tests, that typically remove a large part of the outliers. In the orientation test (left), we see that all matches in the live video image (bottom) are rotated about 45°–60° counterclockwise, compared to the reference image (top). Only the match for feature C is rotated differently, and is therefore considered an outlier and removed. In the geometric test (right), we consider the line between feature A and feature D, and detect that the match for feature E in the live video image (bottom) lies on the wrong side of such line, compared to the reference image (top). We thus consider such match an outlier and remove it.

orientations to absolute rotation using the feature orientations in the database. Since the tracker is limited to planar targets, all features should have roughly the same absolute orientation (Figure 3.4, left). We estimate a main orientation and use it to filter out all features that do not support this hypothesis. Since feature orientations are available, this step is very fast, yet also very efficient in removing most outliers. The second step uses simple geometric tests (Figure 3.4, right): all features are sorted by their matching confidence and, starting with the most confident features, we estimate lines between two of them and we test all other features to lie on the same side of the line both in the camera image and in the database. The third step removes final outliers using homographies, in a RANSAC fashion [34]. Since most outliers have been already removed, this last step typically requires only a few iterations.

Pose estimation. We finally estimate a 6 DOF camera pose from the point correspondences between the feature points observed in the camera image and the original feature points in the database. We use a Gauss-Newton iteration scheme to minimise the reprojection error under a standard camera model.

3.1.1.2 Offline acquisition of the tracking target

PhonySIFT requires a feature database to be prepared beforehand for each tracking target. Our tracking system is limited to planar targets, therefore a single orthographic image of the tracking target is sufficient. Data acquisition starts by building an image pyramid, similar to the realtime pipeline, each level scaled down with a factor of $1/\sqrt{2}$ from the previous one. Features are detected and described as in the real-time pipeline (see Section 3.1.1.1). Finally, all descriptors and their position on the target are stored in a database.

3.1.2 PatchTracker

Unlike PhonySIFT, PatchTracker does not require an offline preparation of the tracking target, but only requires an orthographic image of the tracking target at runtime. This is typically a scaled-down version of the image used for the offline phase of PhonySIFT (Section 3.1.1.2).

PatchTracker uses active search, based on a motion model that predicts where to find features and how they will appear, and only searches for such features within a small area around their predicted location. Such an approach is efficient because it makes use of the fact that typically both the scene and the camera pose change only slightly between two successive frames – the coarse position of features can be therefore predicted from the previous pose estimates. PatchTracker combines three steps (Figure 3.1): prediction, active search and finally pose estimation. Ultimately, the motion model is also updated with the new pose estimate. In the following, we briefly explain all steps.

Prediction. For each new frame, PatchTracker predicts the location where features should appear in the camera image. Feature locations are calculated by projecting the features of the reference image into the camera image, using a camera pose predicted by a motion model.

Active search. After the feature locations have been predicted, we search the features within a predefined search window around the predicted location (Figure 3.5). We first create an affinely warped representation of each feature – we use the reference image as a source for the pixel values. The exact location of the feature is then estimated using Normalised Cross-Correlation (NCC) over the search window and, once a good match is found, performing a quadratic fit into the NCC responses of the neighbouring pixels, to

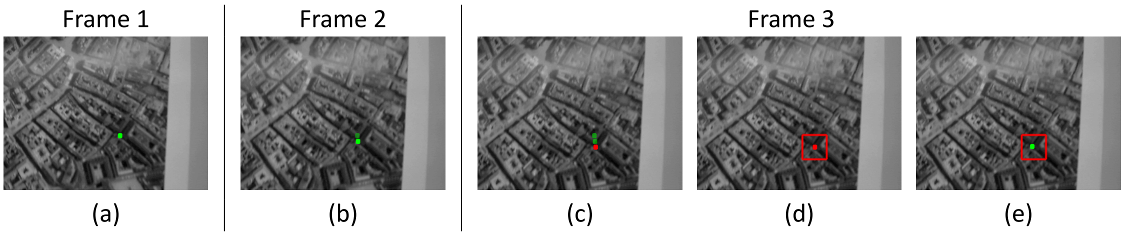


Figure 3.5: Active search in PatchTracker. After detecting the same feature in two consecutive frames (a-–b), we use a linear model to predict where the feature will be located in the third frame (c, prediction shown in red). We then define a search region around the predicted feature location (d). Typically, the actual feature is found within such search region (e, shown in green).

achieve sub-pixel accuracy. Since a small search window limits the speed at which camera motion can be detected, we employ a multi-scale approach to track fast moving cameras. First, we estimate the new pose from a camera image downscaled to half size, with a large search radius. If a new pose is found, we refine it using the full-resolution camera image and a smaller search radius.

Pose estimation. The pose estimator is analogous to the one used by PhonySIFT (Section 3.1.1.1). We feed it with the correspondences between the features in the reference image and the features found by the active search.

Update of the motion model. We finally update the motion model with the newly estimated pose. Since our motion model is linear, we use the difference between the current pose and the previous one as a measure for camera velocity. This model works well, as long as the camera motion does not change drastically, which is rarely the case.

3.2 Evaluation

We evaluated both the robustness and the performance of our tracking system with a number of tests, presented in the following of this section.

3.2.1 Tracking targets

We tested the robustness of PhonySIFT on seven different tracking targets (see Figure 3.6), in stand-alone mode as well as in combination with PatchTracker. The targets were selected to cover a range of different objects that might be of interest in real applications, to gain understanding on what are the characteristics of a good tracking target. We applied our trackers to one test video sequence for each target and measured the number

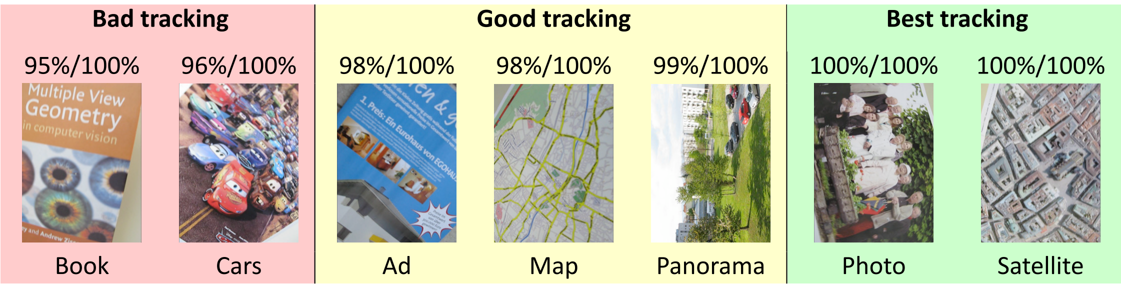


Figure 3.6: Robustness of our tracking system on different tracking targets (we also show the percentage of successfully tracked frames without/with PatchTracker). We ran our tests over seven different tracking targets, covering a broad range of possible real-world AR applications.

of frames for which the pose was estimated successfully (we define a pose to be found successfully, if the number of inliers in the correspondences is 8 or greater). The results are shown in percentage in Figure 3.6.

The Book and Cars datasets performed worst. The Book cover consists of few large characters and a low-contrast blurred image, making it hard for the feature detector to find features in large areas of the target. In the Cars dataset the sky and road are of low contrast and therefore also respond badly to corner detection. Like the Book dataset, these areas are hard to track with our current approaches. The Ad, Map and Panorama datasets show better suitability for tracking. Both the Ad and the Panorama consist of areas with few features, but they are better distributed over the whole target than in the Cars or Book targets. The Map target clearly has well distributed features, but robustness suffers from the high frequency of these features, which creates problems when searching at multiple scales. The Photo and Satellite datasets work noticeably better than the other targets, because the features are well distributed, of high contrast and more unique than the features of the other datasets.

We therefore conclude that drawings and text are less suitable for our tracking approaches. They suffer from high frequencies, repetitive features and typically few colours (shades). Probably, a contour-based approach is more suitable in such cases. Real objects or photos on the other hand have often features that are more distinct, but can suffer from poorly distributed features creating areas that are hard to track.

3.2.2 Overall robustness

We further tested the robustness of our tracking system on five different test sequences, each showcasing a different practical situation that can occur in any real-world AR application (see Figure 3.7). The first sequence resembles a smooth camera path, always

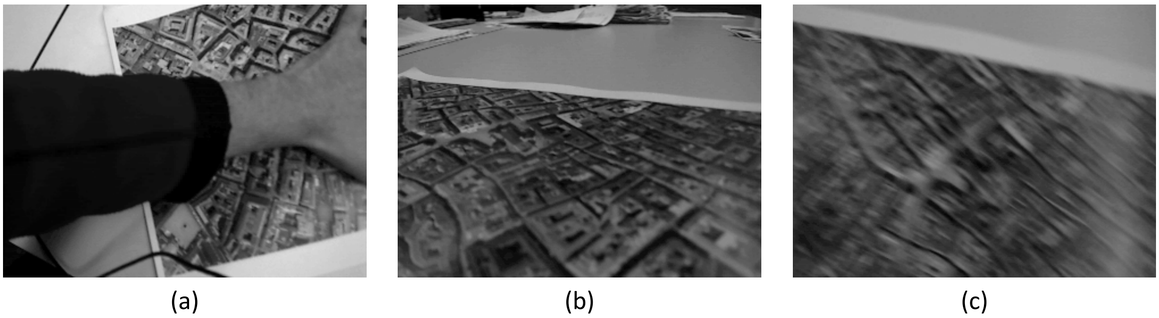


Figure 3.7: We evaluated the robustness of our tracker in various practical scenarios: (a) occlusion, (b) strong tilt, and (c) fast camera movements.

pointing at the target. The second sequence tests partial occlusion due to a user interacting with the target, up to the point that less than one fourth of the target is visible (Figure 3.7 (a)). The third sequence checks how well the tracker works under strong tilt angles, up to about 60° degrees from the perpendicular to the tracking target (Figure 3.7 (b)). The fourth sequence imitates fast camera movement, typical in mobile usage, up to the point of introducing very strong motion blur (Figure 3.7 (c)). The final sequence checks how well the trackers cope with pointing the camera away from the target and back. We used the Satellite dataset to generate these sequences, as it is the dataset that worked best in our previous test.

All five sequences were tested with PhonySIFT alone and in combination with PatchTracker. The results of all tests are shown in Figure 3.8. For each sequence and tracker, we visualise the tracking success (again defined as finding at least 8 inliers) as a horizontal line. The line is broken at those points in the sequence where tracking fails. Tracking success is also shown as the percentage of successfully tracked frames in each sequence, next to the corresponding line in Figure 3.8.

Both trackers work very well with the simple sequence: while PhonySIFT loses tracking for a few frames during the sequence, PatchTracker takes over after the first frame and never loses it. The two variants perform most differently at the occlusion sequence, where large parts of the tracking target are covered by the user's hand. Here, PhonySIFT breaks, while PatchTracker again takes over after the first frame and does not lose track over the complete sequence. PhonySIFT is known to have problems with strong tilts, because it is designed to tolerate tilt but not actively take it into account. Since PatchTracker directly copes with tilt, it does not run into any problems with this sequence. The fast camera movements and hence strong motion blur in the fourth sequence create a severe problem for the FAST corner tracker used for PhonySIFT, while PatchTracker has no problems

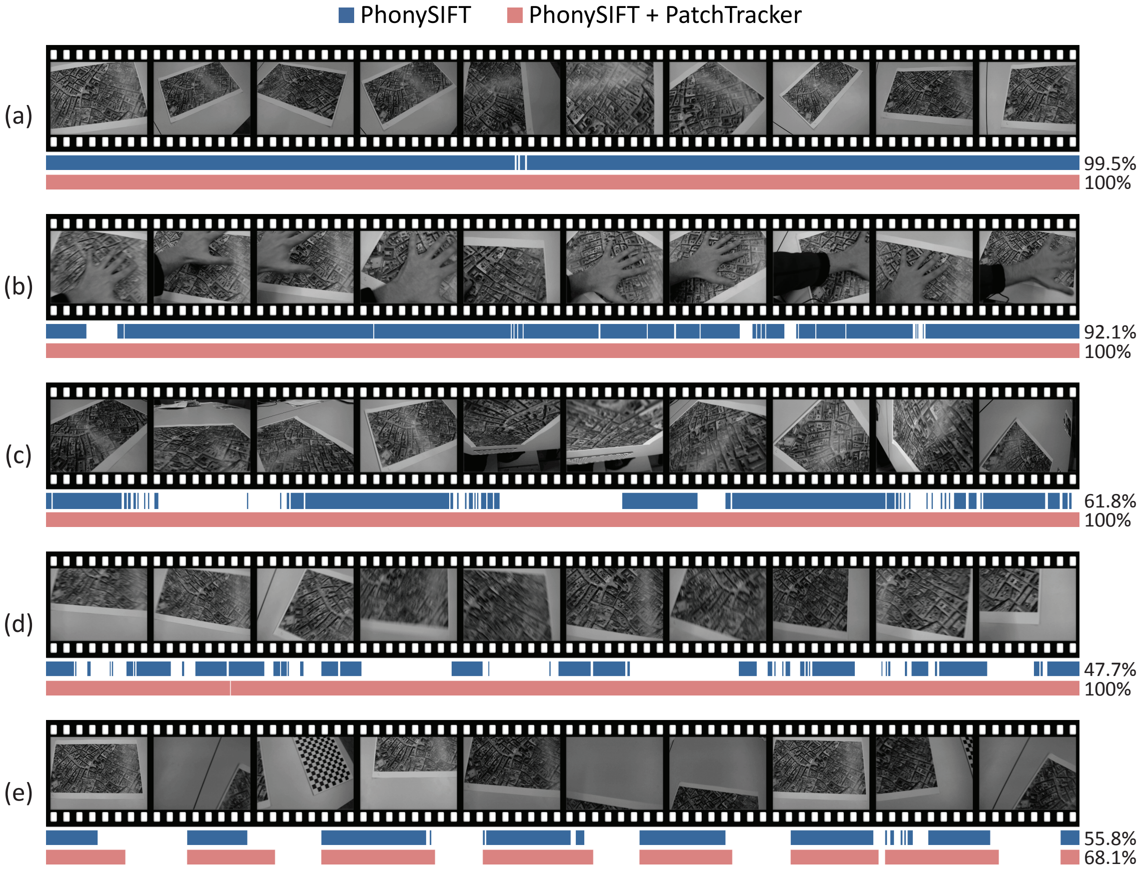


Figure 3.8: Robustness of our tracking system in five practical scenarios. We designed three test sequences: (a) simple, (b) occlusion, (c) tilt, (d) fast movement, and (e) loss of target. The horizontal bars encode tracking success over time, defined as estimating a pose from at least eight keypoints (we also show the percentage of successfully tracked frames). The reference image and test sequences can be downloaded from `http://studierstube.org/handheld_ar/vienna_dataset`.

even with strong blur. The last sequence tests reinitialisation after moving the target out of the camera view. We see that PatchTracker loses the target later than PhonySIFT, while the combined PhonySIFT+PatchTracker clearly reinitialises at the same time as the standalone PhonySIFT tracker (because in both conditions the reinitialisation is based on PhonySIFT).

We also looked at the reprojection error of our trackers for all sequences. The accuracy results, presented in more detail in our publication [202], show that both our trackers operate on average with sub-pixel accuracy.

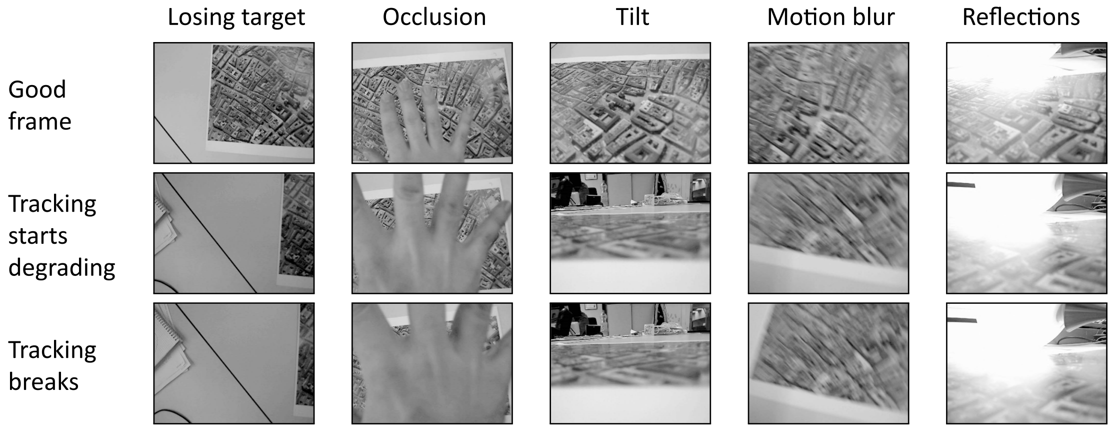


Figure 3.9: Robustness of PatchTracker. We tested PatchTracker against losing target, occlusion, tilt, motion blur, and reflections. The first image of each column shows a typical frame of the test sequence that was tracked well. The second image shows when the tracking quality starts to degrade. The third image shows the first frame that breaks tracking.

3.2.3 Robustness of PatchTracker

The robustness tests in the previous section show that PatchTracker strongly improves the overall tracking robustness of our tracking system, resulting in 100% tracked frames in most of our tests (see Figure 3.8). We therefore created an extra series of tests, in which each image sequence makes the tracking increasingly more difficult, allowing us to analyse when exactly PatchTracker breaks. All sequences were designed so that PhonySIFT detects the pose in the first frame and then hands over tracking to PatchTracker, until it loses the target.

Figure 3.9 shows a typical good frame for each test, the first frame when tracking degrades and the first frame, in which tracking breaks. As PatchTracker was configured to search 100 features per frame, we defined degrading as finding less than 100 features.

The first sequence tests at which point the tracker loses a target moving out of view. Tracking starts degrading when the target covers only ∼17% of the camera image and breaks when it goes down to ∼8%. This percentage clearly depends on the distribution of features on the tracking target. Since the target we used has almost uniformly distributed features (see Figure 3.2), we can assume that PatchTracker works provided that about 8% (or more) features are visible. The second sequence tests partial occlusion of the tracking target. As can be seen in the second image, tracking works well even under large occlusions, and breaks only when the target is hardly visible anymore; as shown in Figure 3.9, second column, the target occupies less than 5% of the pixels in the video image when tracking fails. The tilt test checks how strong the camera can be tilted with respect to the tracking target. Due to the affine warping of patches, very strong tilts can be tolerated and tracking works without degradation, until it suddenly breaks at an angle very close to 90° from the perpendicular to the target (see Figure 3.9, third column). The motion blur sequence tests how fast the camera can be moved – long exposure times are a typical problem of mobile phones, which are not very tolerant to low-lighting conditions. Tracking degrades only when there is severe blur and breaks at a point when the target can be hardly recognised anymore (Figure 3.9, fourth column). Finally, we estimated how well the tracker can cope with reflections on the target. The second image in the last column of Figure 3.9 shows that tracking starts degrading at a point where the camera image is already poor in contrast and then loses the target quickly.

3.2.4 Performance

Finally, since we need to run our tracking system in interactive prototypes, we conducted a further test to verify if it can guarantee sufficient performance on a mobile phone. Performance typically scales linearly with the CPU clock rate on mobile phones, and is independent of the operating system: we therefore benchmarked on a single mobile phone, an Asus P552W, which has a Marvell PXA930 CPU running at 624MHz. We tested the trackers against the simple sequence of the robustness tests of Section 3.2.2, since we wanted to prevent tracking failures and to measure only full-frame processing times. The mobile phone runs PhonySIFT in roughly 40ms per frame. Adding typical overhead of AR applications (camera image retrieval, rendering, application overhead) this performance allows running an AR application at about 15 frames per second on the 624MHz phone of our test. When activating PatchTracker, per-frame time decreases to about 8ms, reducing performance requirements to a level that even average smartphones are capable of tracking at interactive frame rates. We further looked at the relative speed of the different steps of the PhonySIFT and the PatchTracker pipelines: results are presented in Figure 3.10 and in Figure 3.11. Although we did not formally benchmark the tracker on more recent devices, we use it constantly in our research prototypes and we have empirical evidence that tracking runs at full 30fps on the >1GHz CPUs of latest smartphones.

3.3 Conclusion

This chapter presented a natural-feature tracking system that allows robust pose estimation from planar targets at real-time frame rates on mobile phones. We combine simplified

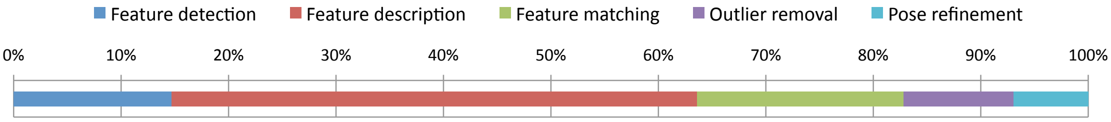


Figure 3.10: Speed of the different parts of the PhonySIFT pipeline.

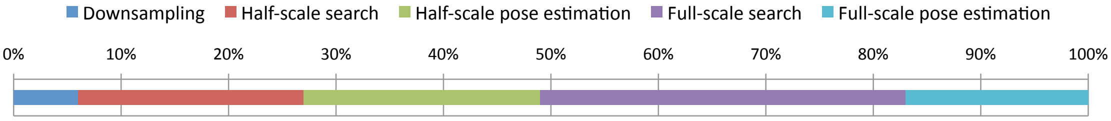


Figure 3.11: Speed of the different parts of the PatchTracker pipeline.

feature detection, description and matching steps with fast but effective heuristics for filtering outlying correspondences. Our robustness and performance tests clearly show the suitability of PhonySIFT for being used in interactive systems.

We further improve the robustness and the performance of our tracking system by carefully considering the usage scenario in real-world settings. By observing that users mostly perform small and smooth movements of the device while interacting with the AR system, we combine PhonySIFT with PatchTracker, a tracker designed to perform fast active-search tracking. This drastically reduces the execution time of our tracker (from 40ms per frame with PhonySIFT to 8ms for PatchTracker, on our test device) and increases the robustness to occlusions, tilts, fast camera movements and strong reflections.

This technical contribution has a large impact on the usability of handheld AR navigation systems. First, systems that use our tracker have a good amount of free CPU cycles available for non-tracking tasks, such as path planning and rendering, while still guaranteeing a smooth interactive experience at 30 frames per second. Second, they can augment real-world planar objects (such as maps or posters) with AR navigational cues robustly, typically without tracking failures that could hinder the user experience.

As we will show in the next chapters of this thesis, this tracking system allowed us to implement an outdoor navigation system based on augmented paper maps (Chapter 6), and an indoor navigation system based on augmented posters (Chapter 8). Our evaluations also informed us while choosing appropriate tracking targets – in Chapter 6, for example, we use a satellite image with street overlay, rather than a pure map, as our work shows that this is a more robust tracking target. Finally, the robustness of our tracker allowed us to successfully conduct a number of evaluations of our prototypes with end users under real-world conditions.

Chapter 4

Tracking panoramas

The previous chapter presented a natural-feature tracking system that enhances augmenting planar targets – such as maps and posters – with AR navigational cues. Such technology does not enable tracking generic three-dimensional environments: this limits the possibilities that can be explored in our research on handheld navigation systems – ideally, we would like to place AR navigational cues not only on maps and posters, but also directly on buildings, streets, and other generic environmental features.

In this chapter, we present another tracking system, designed with the goal of augmenting the physical environment surrounding the user. This second system is based on panorama tracking. Six DOF tracking on mobile phones in outdoor scenarios is still largely an unsolved problem, in particular with the robustness necessary when the tracker must be operated by end users. In our panorama tracking system, we constrain the problem to 3 DOF, assuming that the user will operate the system while standing still and only perform rotational movements. This assumption is often reasonable when users are focused on exploring the AR cues. The assumption of pure rotational movements is typically not respected in real-world usage: users do not rotate their device on the camera axis but rather with extended arms, for more comfort. However, in most outdoor scenarios the distance between the camera and the objects in the environment is large, compared to the translational motion that occurs when users rotate the handheld device. DiVerdi et al. [30] show that, in this case, errors in the panoramic mapping process are negligible.

Due to this simplified 3 DOF problem, our system allows for robust real-time orientation tracking on mobile phones, using solely the computational resources and the built-in camera of the device. We use a panoramic map and, optionally, the sensors to track the user's orientation in the environment. In the following section, we first present the vision-based implementation of such panorama tracker. In Section 4.2, we then discuss

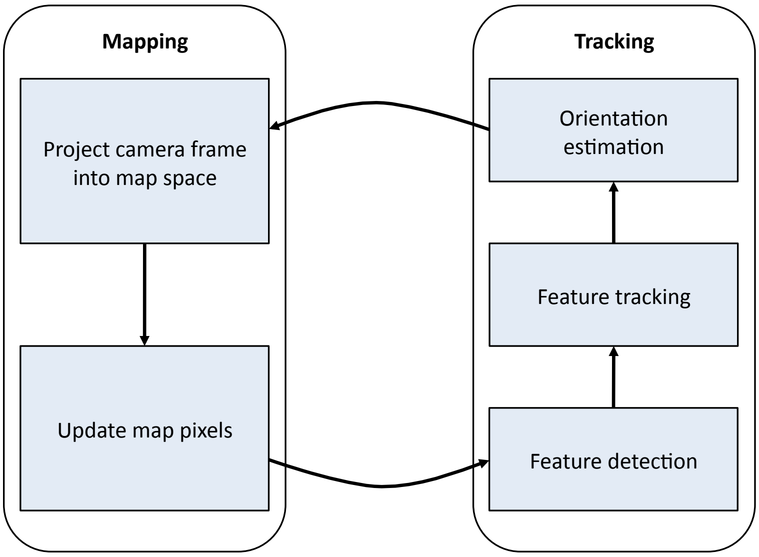


Figure 4.1: The pipeline of our panorama tracking system. We combine two interdependent phases of mapping and tracking.

how we combine our vision-based tracker with sensor measurements, to be able to track the user's orientation in a world-aligned reference frame. Similar to the previous chapter, we give here an overview of our tracking systems, focusing in particular on their impact on handheld navigation systems. For further details on the tracking technology, we refer the reader to our original publications [173, 200].

4.1 Panorama tracker

In this first part of the chapter, we present our vision-based panorama tracking system. Similarly to the tracking system presented in the previous chapter, we rely on natural features for creating the panorama and for tracking the user's movements in the environment. We also present an evaluation of our tracking system, discussing both its robustness and its performance on mobile devices.

4.1.1 Tracking system

Our tracking system works by incrementally creating a panoramic map on the fly from the live camera video, and simultaneously tracking the camera orientation from the same panoramic map.

Figure 4.1 shows how our tracking system is composed of two phases of *mapping* and *tracking*, which depend on each other in our tracking pipeline: while tracking requires the

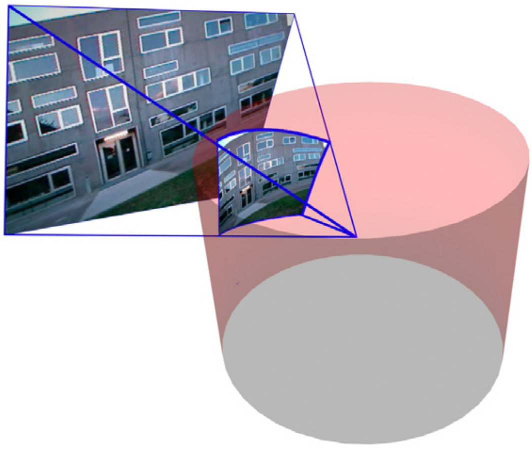


Figure 4.2: Mapping of the camera image onto the cylindrical panoramic map. The mapping process is composed of two consecutive steps: first, we determine the area on the cylinder which is covered by the current camera image. Then, we fill all pixels of such an area with the pixel colours from the camera image.

panoramic map for estimating the orientation of the user's device, mapping requires an orientation from the tracker for updating the map with the data from new camera images.

4.1.1.1 Mapping phase

The mapping phase of our tracker uses a cylindrical map to capture and store the panorama on the fly. Our panoramic map covers 360° horizontally, while the range covered vertically is [-38.15°, 38.15°]. We consider this a reasonable tracking range for typical urban scenarios. Our system generates a colour panoramic map of 2048 × 512 pixels, which can also be used for user-interface purposes.

For each camera image, we assume that the camera position is in the centre of our mapping cylinder and project the camera frame onto the cylinder, to identify which areas of the panoramic map are covered by the current frame (see Figure 4.2). Once we have identified such areas, we conduct a second step to fill the panoramic map with the pixel values from the camera image. In the following, we explain both steps.

Projecting from camera image onto the map. We first define a bounding rectangle for the whole camera image in camera space, setting its vertices as the pixel coordinates of the corners of the camera image. Given the current camera orientation, we then use *forward mapping* to estimate the projected position of the vertices of the bounding rectangle onto the cylinder's surface. More details on the equations used in forward mapping can be found in our paper [200]. Due to radial distortion and the nonlinearity of the mapping, we sample five points on each side of the rectangle – rather than just the two extremes –

to get a smooth curve in the target space.

Filling the map with pixels. This forward-mapped bounding rectangle tells us which pixels in the panoramic map are covered by the current camera image. Since using forward mapping to fill the map with pixels can cause holes or overdrawing of pixels, we rather use *backward mapping* for this operation. We again refer the reader to our paper [200] for the equations used in backward mapping. Essentially, for each of the pixels in the panoramic map identified in the previous step, we apply the inverse of the transformation used for forward mapping, to calculate the corresponding pixel coordinate on the camera image. The resulting coordinate generally lies somewhere between pixels, so we finally interpolate linearly to achieve a colour with sub-pixel accuracy. Since a complete backward mapping of all camera images would cause a large computational workload, we set each map pixel only once, as soon as it can be mapped for the first time.

Loop closing. Due to errors that accumulate as the panoramic map extends away from the starting orientation, after a full 360° sweep the edges of the panorama typically do not align correctly. Rather, there is often a noticeable discontinuity at the location in the map where the left-most and right-most mapped pixels touch. Our system integrates a loop-closing procedure (see Figure 4.3) that accurately estimates this error in the map and corrects it. We explain this method in detail in our paper [200].

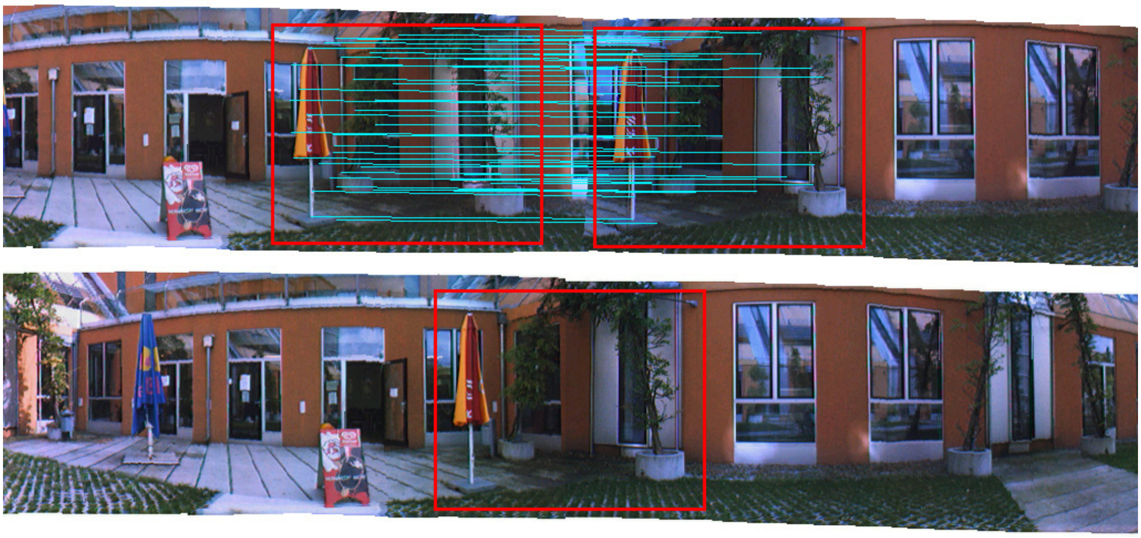


Figure 4.3: The typical output of our panorama tracker before and after loop closing. Top: a full 360° loop with overlapping areas highlighted. Bottom: the same panorama after the application of our loop-closing procedure.

4.1.1.2 Tracking phase

The previous section describes how we build and store a panoramic map on the fly, using the live video images from the camera. The whole mapping process assumes that an accurate estimate of the camera orientation is available. In this section, we present an efficient method that uses the panoramic map (as it is being built) for tracking the camera orientation.

Feature detection. Similarly to the feature detector presented in Section 3.1.1.1, our panorama tracker also applies the FAST corner detector [167] on the panorama to extract features. We extract features for the full-scale panorama as well as for downscaled versions at half and quarter resolution. For each feature, FAST also gives a score of how strong the corner appears. We sort all features by corner strength and only keep the strongest features. In a 64×64 pixel area of the panorama we typically keep only the strongest 40 features at full scale, 20 at half scale, and 15 at quarter scale. We empirically found these thresholds to allow for robust tracking with low memory requirements.

Feature tracking. Feature tracking works similarly to the PatchTracker tracker presented in Section 3.1.2. We use the panoramic map as a natural-feature tracking target, and adopt an active-search approach, based on a motion model, to track features from one frame to the next. Due to the active-search approach, the tracker needs a rough prediction of the camera orientation for every new frame, in order to accurately estimate the actual orientation. In the first frame, this corresponds to the orientation used for initialising the system. For all successive frames, we use a motion model with constant rotational velocity (under the assumption of zero translation) to predict the orientation in the next frame. We calculate velocity as the difference in orientation between the current and the previous frame. We then refine this initial prediction to an accurate orientation estimate: for each feature, the tracker applies NCC over a search area at the expected feature location in a camera image scaled down to a quarter resolution, and then refines the feature position to sub-pixel accuracy by fitting a 2D quadratic term to the matching scores of the 3×3 neighbourhood.

Orientation estimation. Feature tracking returns a first set of correspondences between 3D cylinder coordinates (from the panoramic map) and 2D camera coordinates (from the current camera image). We use these correspondences in a non-linear refinement process, using the initial orientation prediction as a starting point. The refinement uses a Gauss-Newton iteration, running the same optimisation as used for a 6 DOF camera pose (see

Section 3.1.1.1) but ignoring all translational terms. Like in Section 3.1.2, we first refine the orientation at half resolution and finally at full resolution, to cope with large camera movements without using large search areas.

Relocalisation. The active-search approach can only follow the features from one frame to the next, and is therefore not able to reinitialise when frame-to-frame tracking breaks. A relocalisation mechanism is fundamental for any practical system. We therefore add a separate relocalisation step, which is triggered if the tracker does not find enough feature matches (we typically set the minimum number of matches to 40 for full resolution, 20 for half resolution and 12 for quarter resolution), or when the reprojection error after refinement is too large to trust the orientation estimate (we typically allow for a maximum reprojection error of 4 pixels). Relocalisation works by storing low-resolution keyframes with their respective camera orientation, as the map is being created. To limit the memory needed to store the keyframes, we downsample the camera image to quarter resolution. The orientation is then quantised into 12 bins for yaw (±180°), 4 bins for pitch (±30°) and 6 bins for roll (±90°): we store a unique keyframe for each bin (we thus have a maximum of 288 keyframes). A keyframe is only overwritten if it is older than 20 seconds. Whenever active-search tracking is lost, we use brute-force NCC to compare the current camera image with all the stored keyframes. If we find a keyframe that matches the current camera image, we run the orientation-refinement step using the keyframe's orientation as a starting point.

4.1.1.3 Initialising from a previous panorama

One advantage of our panoramic tracking is that it works without any previous knowledge of the environment, because the panoramic map is not prerecorded but created on the fly. If an application needs to use the tracker on prerecorded panoramas, for example to completely avoid mapping errors, we propose another method that allows initialisation from a prerecorded (partially) finished panoramic map (Figure 4.4).

Starting with a map loaded from a file, we extract features from the map and create PhonySIFT descriptors (Section 3.1.1.1). We then match all features in the camera image against all features in the panoramic map, obtaining a set of correspondences. We finally apply a RANSAC approach [34] to estimate the camera orientation from the list of correspondences. As soon as initialisation successfully completes, tracking continues as described in the previous part of the chapter.

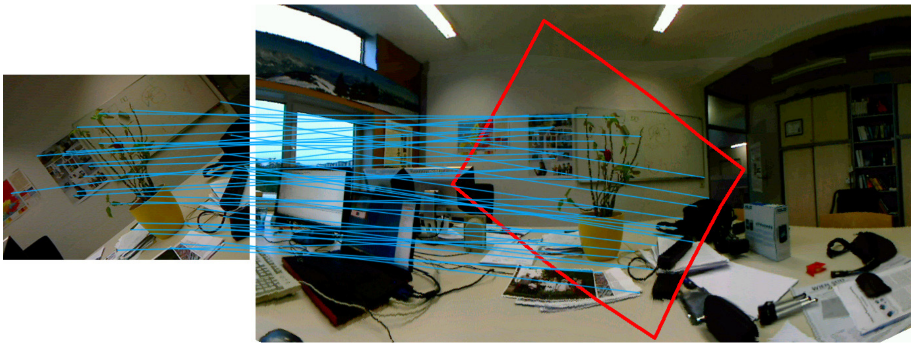


Figure 4.4: Initialization of the panorama tracker from a prerecorded panorama. Left: camera image to localize. Right: 3DOF localization of the camera image on the prerecorded panorama.

4.1.2 Evaluation

We conducted an evaluation of our panorama tracker, to analyse its robustness, accuracy, and to verify that it allows interactive usage on a mobile phone. To evaluate our panorama tracker, we created 30 panoramas at different indoor and outdoor locations.

4.1.2.1 Robustness

We specifically included difficult scenarios, to investigate when the tracker breaks. For 5 of the 30 panoramas, tracking broke before we could finish a 360° loop.

In Figure 4.5, we qualitatively analyse where tracking breaks, showing a few practical examples. We can see that the most typical problems are poorly textured surfaces or surfaces with unevenly distributed features, and also moving objects in the environment. This knowledge is useful to inform our experimental designs, suggesting that we should avoid poorly textured and extremely crowded environments for our evaluations.

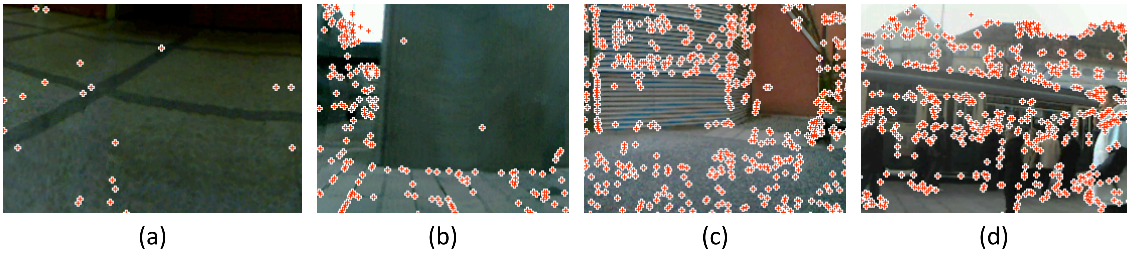


Figure 4.5: Problematic scenarios for our panorama tracker. A few cases are typically problematic: (a) a floor that is poor on texture, (b) a wall that is poor on texture, with only line details on the floor, (c) repetitive line details on the wall (features are due to sampling artifacts) and pebbles on the floor, (d) moving objects (tram, people) covering most of the image.

4.1.2.2 Accuracy

We evaluated the accumulated mapping error, using only the 25 complete panoramas, and measured the offset in the overlapping regions using our loop-closing method. With a proper camera calibration, 16 out of the 25 panoramas have an error of ∼1° (given the panorama width of 2048 pixels, this corresponds to an error of about 5–6 pixels). We also analysed pose jitter, using a static live camera (zero motion): we measured a jitter of ∼0.05° for both head, pitch and roll angles (this corresponds to a fraction of a pixel). These results show that our panorama tracker is sufficiently accurate for adoption in handheld navigation systems.

4.1.2.3 Performance

We finally benchmarked the performance of our panorama tracker on an Asus P565 smartphone with an XScale ARM CPU at 800MHz. Overall, the panorama tracker takes an average of ∼15 milliseconds for each video frame – about 10.6 milliseconds for the tracking phase and 4.5 milliseconds for the mapping phase. Timings further decrease once the map has been completely filled, or if the user points the camera only at areas that have already been mapped, because the mapping time goes close to zero in such cases. Although we did not formally benchmark the tracker on more recent devices, we use it constantly in our research prototypes and we have empirical evidence that tracking runs at full 30fps on the >1GHz CPUs of latest smartphones.

Loop closing is a more expensive operation, taking ∼10 seconds for a full colour map on the 800MHz phone, when using the Lanczos filter for pixel sampling. Most of the time is spent on updating all pixels in the map, thus loop closing takes less than 2 seconds on the phone if we replace Lanczos filtering with nearest-neighbour filtering. This creates noticeable visual artifacts, but generates a panorama that is still good for tracking. In practice, it is reasonable to assume a small slowdown of the system for loop-closing, because it is a rare situation and is required at most once for a map. Localisation from a pre-recorded panoramic map takes ∼120 milliseconds per frame, and can therefore run interactively.

4.1.3 Discussion

The panorama tracker presented in this section allows for accurate, robust and drift-free orientation tracking and is sufficiently efficient to run at high frame rates on common smartphones. As a side effect of the tracking-and-mapping approach, the method also creates coloured panoramic maps, which can be used for the user interface (as we do, for example, in the case of Section 5.2.1).

A limitation of our panorama tracker is that it can only estimate a camera orientation *with respect to the initial orientation*. In AR navigation systems, we want to augment the physical environment with virtual cues, and we therefore need tracking of the device orientation *with respect to the world reference frame* (i.e., correctly aligned to gravity and north directions). A naïve way to track orientation in world coordinates is to use the data from the magnetometer and accelerometer to initialise the panorama tracker. However, sensor measurements are typically noisy, particularly on mobile phones. Due to this noise in sensor measurements, directly using them would mean initialising the panorama tracker most of the times with an inaccurate – or completely wrong – world orientation. A more robust approach is to consider all sensor measurements over a large time span, integrating them to estimate a world registration of the panorama tracker that gets incrementally more accurate. We explain this method in the next section.

4.2 World-aligned panorama tracker

In this tracking system, we still use our vision-based panorama tracker (presented in the previous section) as the main tracking modality. At the same time, we use sensor measurements to register the panoramic map to a world-aligned reference frame (based on gravity and north directions), as shown in Figure 4.6.

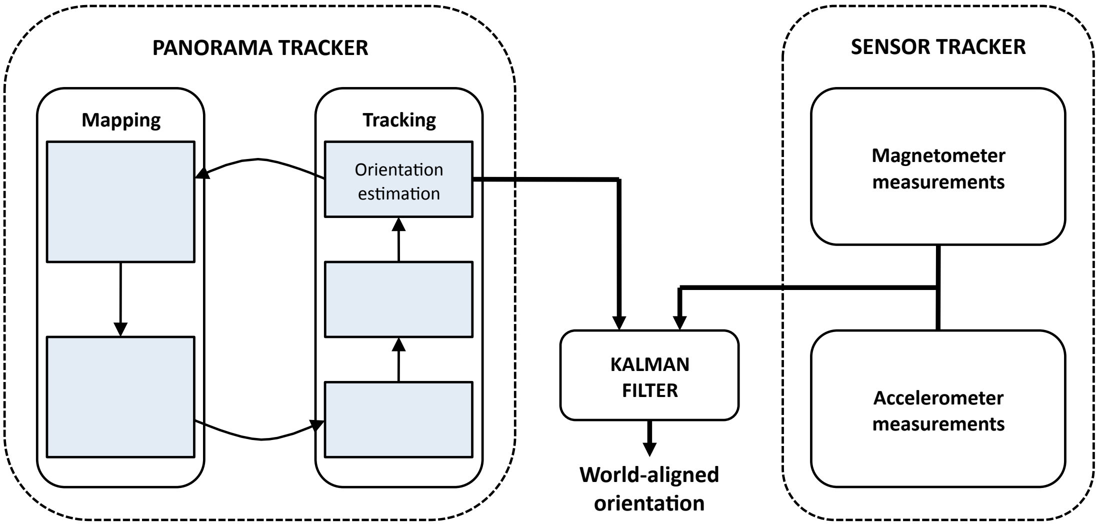


Figure 4.6: Pipeline of our world-aligned panorama tracker. We combine the vision-based panorama tracker (see Figure 4.1) with sensor measurements, through a Kalman filter.

While our vision-based tracker provides accurate and low-jitter orientation estimates with respect to an initial reference frame, sensor measurements provide inaccurate and jittery orientation estimates with respect to a world-aligned reference frame. The outcome of our approach is therefore an accurate and low-jitter orientation estimate with respect to the world reference frame, that can be directly used in handheld navigation systems to place accurate augmentations in the environment.

4.2.1 Tracking system

The key idea behind our tracking system is to continuously refine an estimate of the rotational offset between the panorama tracker and a sensor-based orientation tracker (see Figure 4.7). The sensor-based tracker assumes a world-aligned reference frame N, defined by the direction to magnetic north and the gravity vector. The accelerometer and magnetometer on the mobile phone measure the orientation of the device D with respect to this reference frame. The output of the sensors is therefore a rotation R_{DN}[*]. Our panorama tracker, in contrast, provides a rotation of the device R_{DP} with respect to the reference frame P of the panorama tracker.

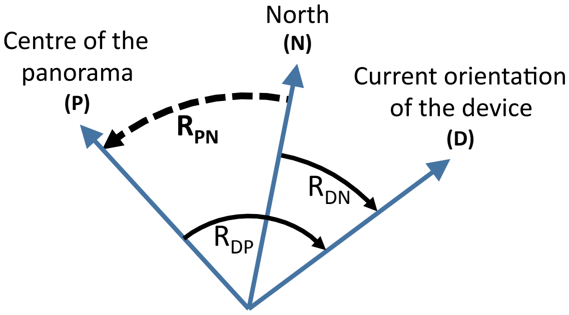


Figure 4.7: Overview of the reference frames involved in our tracker. Different rotations are calculated between a world reference system (N), a device reference system (D) and a panorama reference system (P). We use the two rotation measurements R_{DP} and R_{DN} to estimate the invariant rotation R_{PN}.

Our goal is to estimate the invariant rotational offset R_{PN} in Figure 4.7. At each frame, this rotation can be simply estimated as

$$R_{PN} = R_{DP}^{-1} \cdot R_{DN}$$

We use an extended Kalman filter (EKF) to integrate all estimates of the rotation R_{PN} over time. Essentially, for every frame, the filter estimate's $\hat{R}_{PN}$ and the sensors'

[*] The subscripts in R_{BA} should be read from right to left, to signify a transformation from A to B.

measurement R_{PN} are used to compute a small innovation motion R_i, as shown in Figure 4.8. This innovation is then fed into the usual update equations of the EKF, presented in more detail in our publication [173].

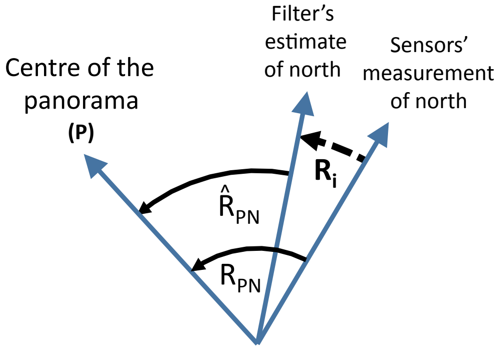


Figure 4.8: The difference between the filter's estimate and the sensors' measurement. We calculate the innovation motion R_i using the Kalman filter's estimate $\hat{R}_t$ and a new sensors' measurement R_{PN}.

The orientation of the device within the world-aligned reference frame is finally computed through concatenation of the estimated orientation $\hat{R}_{PN}$ from the EKF and the measured orientation from the panoramic tracking R_{DP}, as

$$R_{DN} = R_{DP} \cdot \hat{R}_{PN}$$

This combines the accurate orientation from the panorama tracker with a filtered estimate of the world-aligned orientation from the sensor-based tracker.

4.2.2 Evaluation

We evaluated our tracking system on an HTC HD2, a smartphone based on a Qualcomm Snapdragon 1GHz CPU. The phone uses an AKM AK8973 3-axis electronic compass and a Bosch BMA150 3-axis acceleration sensor.

4.2.2.1 Accuracy

We tested the absolute accuracy of our tracker using a set of surveyed reference points, whose position we know with centimetre accuracy. Figure 4.9 shows the position of all reference points (1–8) and the reference point where the phone was positioned (RP). Given the accurate geographic coordinates of all reference points, the absolute angle from the phone's position RP to north of each target point can be calculated geometrically.

We mounted the phone onto a tripod (Figure 4.10, left), positioned exactly above the reference point RP (we used a perpendicular to ensure the correct location of the tripod). We estimate that an error of ±2cm in the placement of the tripod would lead to an orientation error of ±0.02 degrees.

Due to the high sensitivity of the magnetometer to metal parts, we could not use a pan-tilt unit for measuring accuracy. We therefore measured accuracy by manually turning the device towards all reference points, without resetting the tracker. We kept the device still for about 30 seconds at each reference point, logging the orientations reported by the sensors, by the vision tracker and by the combined tracker. A viewfinder glyph on the device's screen (Figure 4.10, right) ensured pixel-accurate alignment of the camera with the target points. Figure 4.11 (left) shows the orientation reported by the sensors and by the combined tracker during this measurement. Reference angles are shown as dark dotted lines. Figure 4.11 (right) shows the error to the closest reference point, effectively the error to the ground truth. We also plot the error of the vision-based tracker as a reference for the hybrid tracker. Since the vision-based tracker does not provide absolute orientation from the north, in the plot we assume it has zero error on the first sample.

The results demonstrate two key improvements over pure sensor-based tracking. First, high frequency noise is reduced: the vision-based tracker dominates the motion estimation and provides a low-jitter orientation estimate. Secondly, over time, the error of the combined tracker is smaller than the sensor-only orientation, because deviations in the magnetometer measurements are averaged over different orientations. Overall, we obtain a responsive, less jittery estimate that, on average, is also more accurate than the orientation derived from the sensors alone.

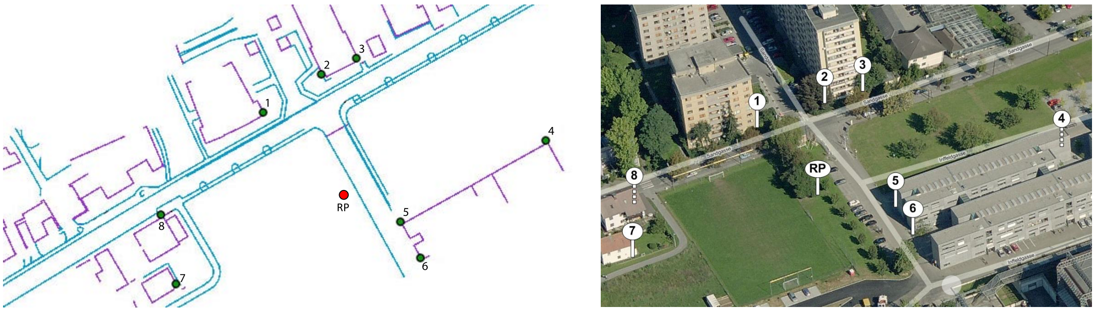
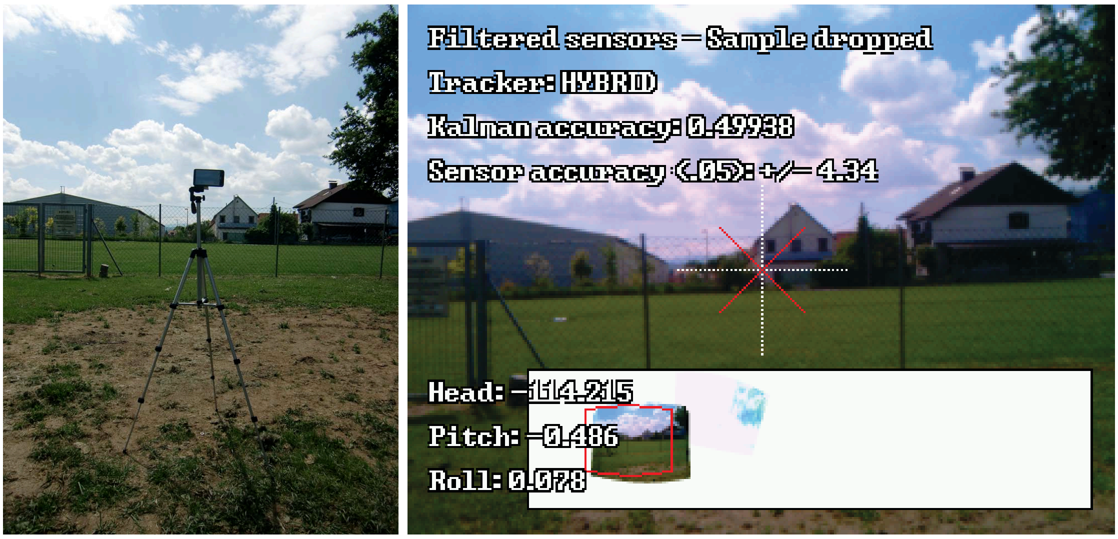


Figure 4.9: The area used for our accuracy evaluation. We know the position of all reference points with centimetre accuracy: (left) north-aligned map and (right) bird's eye view.



Filtered sensors – Sample dropped
Tracker: HYBRID
Kalman accuracy: 0.49938
Sensor accuracy (.05): +/– 4.34
Head: –114.215
Pitch: –0.486
Roll: 0.078

Figure 4.10: Phone-based setup used for our accuracy evaluation. We mounted the device on a tripod (left) and ran an application with a viewfinder (right), to help us align the device with the different reference points.

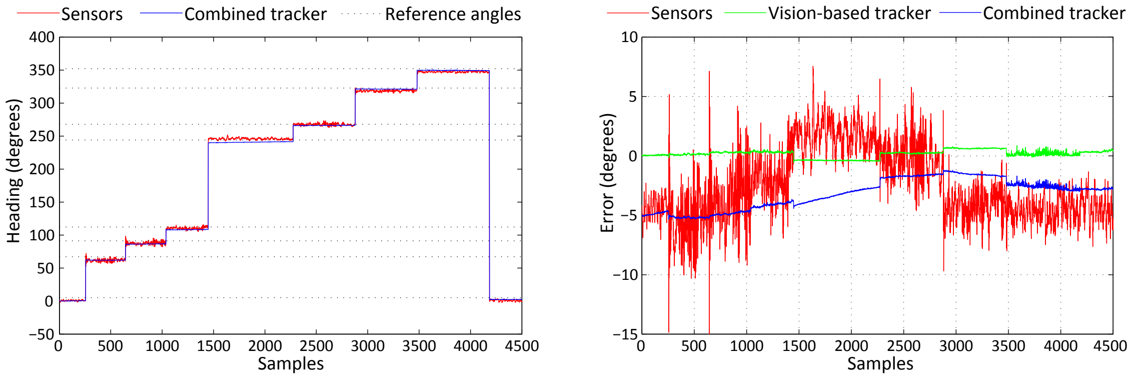


Figure 4.11: Accuracy results for a phone fixed on a tripod. We plot the results of a test sequence for our tracker: (left) angle reported by the sensors and by our orientation tracker, with respect to a set of reference points; (right) angular error of sensors, vision-based tracker and combined tracker.

4.2.2.2 Free-hand motion

The tripod-based accuracy evaluation provides a measure of absolute accuracy of our tracker. We also tested free-hand motion (holding the device in the hand), evaluating the tracker's behaviour in a more realistic usage scenario. Figure 4.12 shows data captured while rotating the phone from one reference point to another (represented by the two dotted lines) through a natural rotational movement – the phone was held in the experimenter's hand. We plot raw sensor measurements, filtered sensor measurements[†] and the output of our tracker. The plot shows that our tracker removes jitter without inducing latency, in contrast to the filtered sensor estimates.

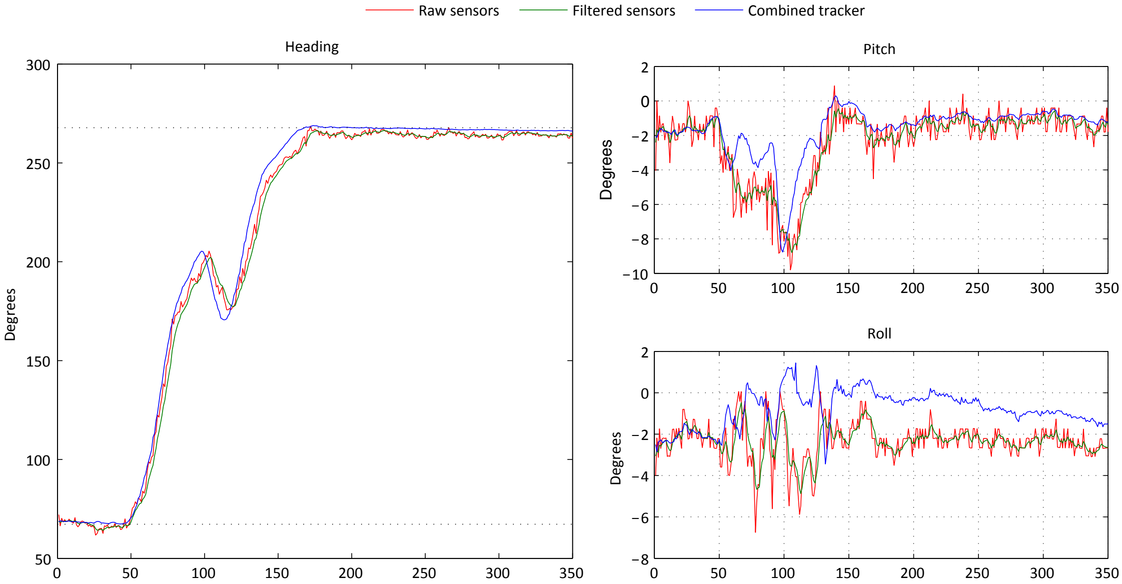


Figure 4.12: Accuracy results for free-hand motion. We plot heading, pitch and roll for a free-hand movement of the mobile phone between two reference points (dark dotted lines). We plot orientation for the raw sensor values, a filtered estimate and the hybrid tracker.

4.2.3 Discussion

The tracking system presented in this section significantly improves the orientation estimation of the sensors built into modern mobile phones. Further, it obtains a world-aligned orientation estimate from the accurate and low-jitter vision-based tracker presented in the first part of this chapter. The implementation as a recursive filter is efficient and fast, requiring only a little more memory and processing power than our original panorama tracker. Therefore, like the panorama tracker alone, it can run at interactive frame rates in handheld AR applications operating on smartphones.

[†] We filter the raw sensor measurements with a Kalman filter tuned for low latency and reasonable filtering of high-frequency noise.

4.3 Conclusion

This chapter presented a panorama tracker that allows robust orientation estimation at real-time frame rates on mobile phones. Similar to the natural-feature tracking system presented in the previous chapter, we model our tracker around observations on how users typically operate the AR system. We first design the tracker around the assumption that users will operate the system while standing still: this reduces the pose-estimation problem from 6 DOF to 3 DOF, allowing us to robustly create the panoramic map on the fly without previous knowledge of the scene. We then observe that users will perform small and smooth movements for a large part of the interaction time: this allows us to apply again an active-search approach, which guarantees high processing speed of our tracker. We finally combine our panorama tracker with sensors, obtaining a fast and low-jitter orientation tracker that returns a pose in a world coordinate frame.

This technical contribution expands the design possibilities that we can explore in handheld AR navigation systems. Systems based on our panorama tracker can augment the physical environment with AR cues guaranteeing a low-jitter and smooth interactive experience at full camera frame rates. As we will show in the next chapters of this thesis, this tracking system allowed us to implement two outdoor navigation systems, one for browsing generic information in the environment (Chapter 5) and another one for guided turn-by-turn wayfinding towards a destination (Chapter 7). As our results will show, the assumption that users will stand still while using the system (and only perform rotational movements) is reasonable when users are exploring information, while it seems too strict for the case of users finding their way in the environment (in this case, a full 6 DOF tracker might be better suited).

PART III

Exploration

Chapter 5

Egocentric exploration

In this part of the thesis, we focus on the task of *exploration* – browsing and inspecting the information in the surroundings, and gaining understanding of its spatial location in the environment. During exploration, AR can support this understanding with *egocentric views*, augmenting the information directly at its corresponding physical location in the environment (for example, a shop's name is augmented onto the corresponding building's façade). Consequently, users can inspect the information just as naturally as they would inspect the physical environment.

When browsing information in an egocentric view, AR intrinsically imposes two constraints: the point of view is bound to the video camera (mounted on the device) and the field of view is limited by the camera optics. This can result in some of the information being off screen and therefore not visible by the user. Furthermore, the limitation on the field of view also prevents the user from gaining a 360° overview on the surrounding information. This chapter investigates *how to support users' understanding of the surrounding information and its location*, within an AR navigation system.

5.1 Pointing to a specific augmentation

This first part of the chapter targets the case in which the system knows that the user is looking for a specific piece of information: the goal of the interface is to make the location of such information clear to the user. We therefore investigate how to point to a specific piece of information within AR, and how to make the physical location of such information clear to the user.

We use the scenario of a user browsing live information from remote cameras in outdoor

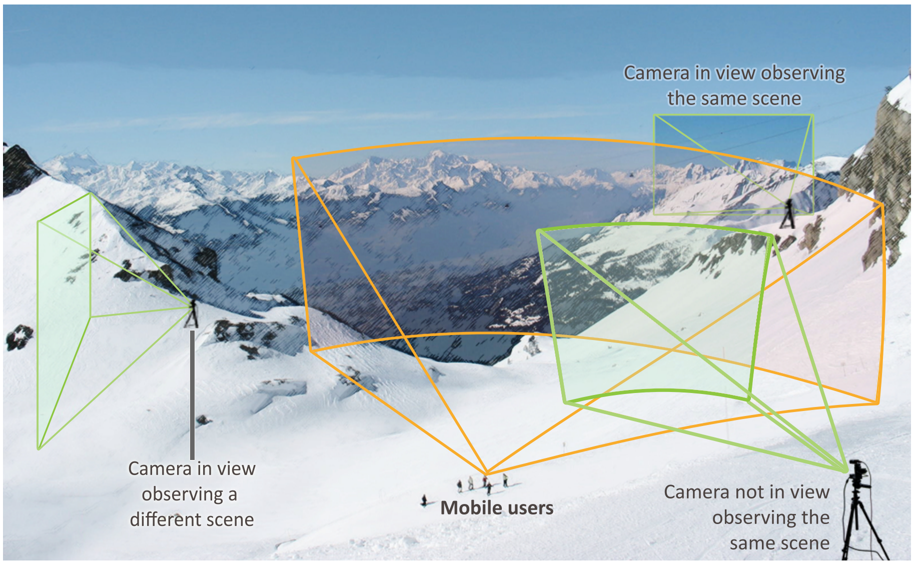


Figure 5.1: A sample scenario for the HYDROSYS project. Mobile users need to access a number of remote-camera feeds and to understand the location of remote cameras with respect to their own position.

settings. This is a typical scenario for the HYDROSYS* project, which is focused on environmental monitoring using mobile devices. In the case of environmental monitoring, remote cameras help users in obtaining an overview of the situation, planning for movement or searching for anomalies (e.g., assessing the risk of avalanches). Teams of users monitor environmental changes on-site using mobile AR.

A typical scenario is shown in Figure 5.1. Remote cameras are placed in “wild” areas such as mountains, riverbanks, ridges, and are accessed by mobile devices through a sensor network. Users often have little knowledge about the surroundings, but they need to understand the position of the remote cameras in order to evaluate the situation. Remote cameras may or may not be visible by the user depending on her position and on occluding objects in the environment. In browsing the camera feeds, navigation cues are necessary to help understand the position and orientation of a remote camera.

5.1.1 Interface design

In our interface, AR is used to render a remote camera as a virtual 3D view frustum and a live video feed anchored to the physical position of the camera in the environment (see Figure 5.2). The frustum matches the position of the camera and the direction it is looking.

*HYDROSYS project: `http://www.hydrosysonline.eu/`

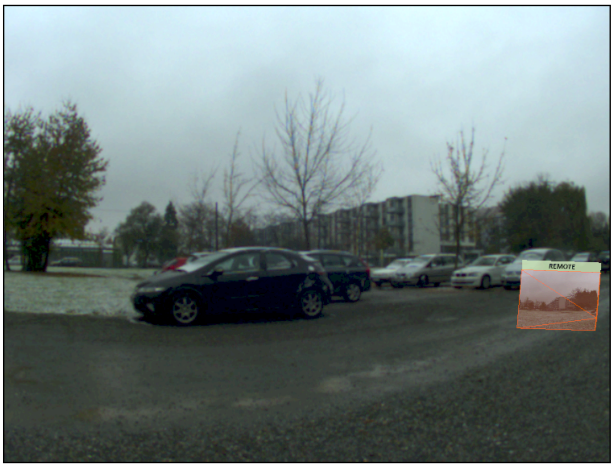


Figure 5.2: Visualisation of a remote camera in our AR view.

We designed three interfaces for pointing to the view frustum and clarifying the spatial relationship between the user and the remote camera (see Figure 5.3). All interfaces operate in three states, shown by the large frames in Figure 5.3. In the first state, the user sees the video feed from the camera mounted on her device (*local camera*) augmented with the 3D frustum of the remote camera (if the remote camera is in view). The view is analogous to the one shown in Figure 5.2. In the second state, a transition view shows both the local and the remote camera feeds, and illustrates their mutual position and orientation. In the third state, the user sees the video feed of the remote camera. The three techniques represent the main research directions on navigation and multi-camera systems, adapted to mobile AR.

Mosaic. Mosaic (Figure 5.4 (a)) relates local and remote cameras topologically using a compass metaphor, through screen-aligned 2D overlay graphics. This technique is typical in surveillance systems, for example in DOTS [42]. The technique uses the angle between the viewing direction of the local camera and the position of the remote camera to position thumbnails of both videos on the screen. This conveys to mobile users how they should turn in order to see the camera, like a compass. Mosaic does not show the 3D spatial relationship or the distance between cameras. It is primarily a 2D technique, providing a directional cue towards the remote camera with respect to the user. The transition view shows thumbnails of both videos, allowing users to get a minimised view of both cameras at the same time. Since the organisation of the thumbnails does not depend on distance, this technique provides a visualisation of several cameras simultaneously, as long as the cameras are not in the same direction.

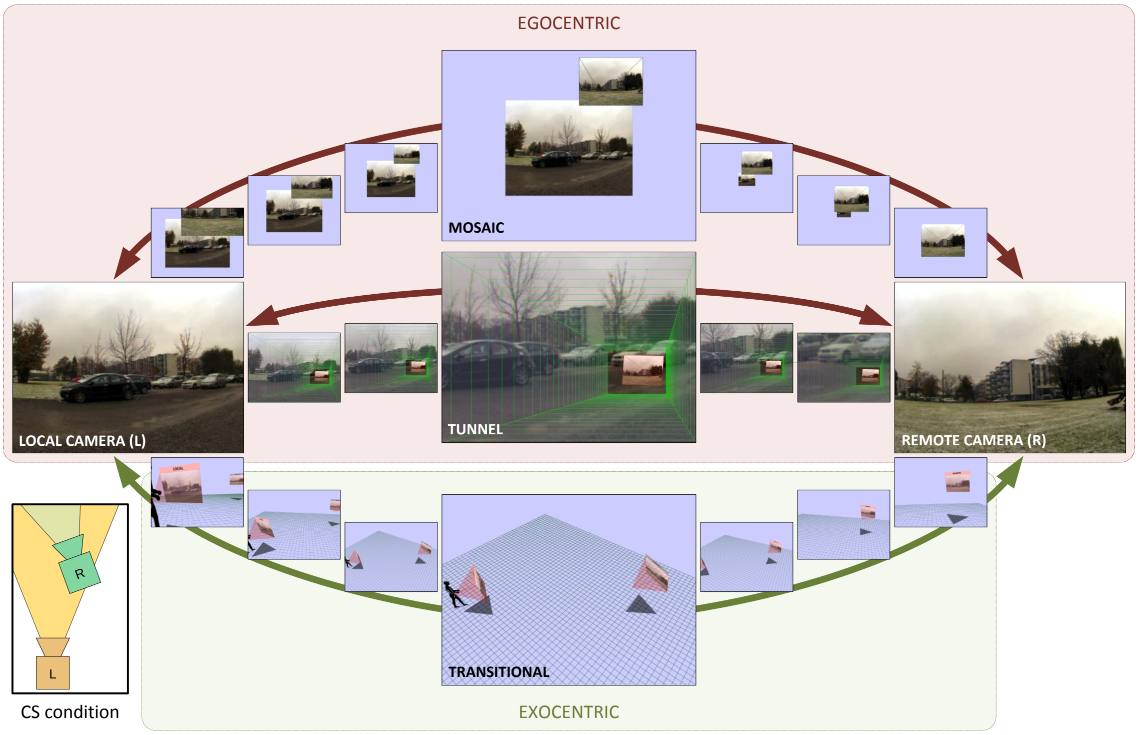


Figure 5.3: An example of the three proposed techniques. Using the techniques, users can browse the video stream from either the local or the remote camera, or they can smoothly move to a transition view where both videos are visible.

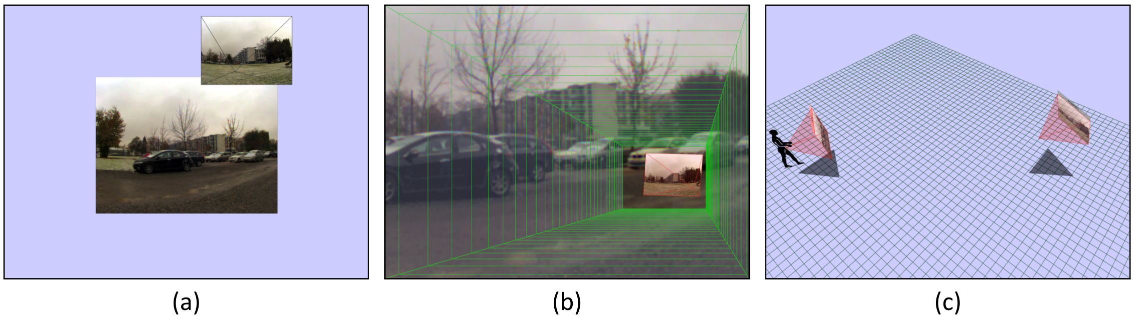


Figure 5.4: Screenshots of the three techniques: (a) Mosaic, (b) Tunnel, (c) Transitional.

Tunnel. Tunnel (Figure 5.4 (b)) is a variation of the attention funnel, first introduced by Biocca et al. [14] and later refined by Schwerdtfeger et al. [175]. This is an AR technique to guide a user towards an object of interest. The technique displays a tunnel, oriented in 3D towards the remote camera. Users can travel down the tunnel to the other camera. We blend the tunnel over the video background so that the tunnel and the video are both visible. When the remote camera is in view, the user can see its video feed at the end of the tunnel. If the remote camera is not in view, the tunnel instructs the user in turning the device towards the camera. Travelling down the tunnel shows the view to the other camera. The distance to the camera is correctly shown in 3D from the perspective of the user.

Transitional. Transitional (Figure 5.4 (c)) implements the concept of a transitional interface as defined by Grasset et al. [45]: in a transitional interface, users can transition between contexts, each possibly having a different space, scale and representation. In our case, users transition between an egocentric AR context, where a full-screen augmented video is visible, and an exocentric 3D virtual context, where users get a bird's eye overview on both cameras and their respective spatial position and orientation. In the exocentric view, an avatar is used to disambiguate the user's camera from the remote camera. We employ smooth animations to support coherent transitions.

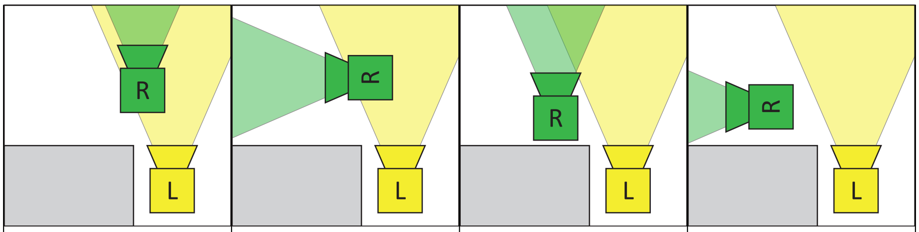
5.1.2 Controlled study

We conducted a comparative evaluation of the three techniques, to gain insight on how effective they are in pointing towards a specific piece of information.

Task. We asked participants to browse a remote camera's video feed, and to understand the position and orientation of the remote camera with respect to their position.

Independent variables. We had two three-way independent variables for this experiment:

- *Technique.* Mosaic, Tunnel or Transitional.
- *Camera configuration.* We classify the spatial relationship between the user and a remote camera by taking into account whether the remote camera is visible from the user's position or not, and also whether the remote camera is observing the same scene as the user, or a different scene. This results in four possible combinations (Figure 5.5): **CS** (camera in view, same scene), **CnS** (camera in view, different scene), **nCS** (camera not in view, same scene), **nCnS** (camera not in view, different scene). In the present experiment, we only consider the first three conditions: in nCnS the spatial relation between the user and remote camera cannot be inferred without prior knowledge from the user (e.g., familiarity with the scene seen by the remote camera). As this condition depends on each participant's expertise, it would have introduced a strong confounding factor in the experiment.



Visibility:	Camera in view	Camera in view	Camera not in view	Camera not in view
Scene:	Same scene	Different scene	Same scene	Different scene
Condition:	**CS**	**CnS**	**nCS**	**nCnS**

Figure 5.5: Different camera configurations. For a local camera (L) and a remote camera (R), we classify camera configurations based on the remote camera's visibility from the user's position, and on the scene the remote camera is looking at.

Research questions. We aimed to answer, for each camera configuration, the following research questions:

Q1. Are there differences in spatial awareness[†] with the different techniques?

Q2. Which technique has less impact on the mental workload[‡]?

Q3. What is the users' subjective preference between the techniques?

Procedure. We conducted the study as a 3 (Technique) × 3 (Camera configuration) design. We treated Camera configuration as a between-subject variable: users were divided in three groups and each group experienced a different type of camera configuration. Technique was treated as a within-subject variable. Hence, every participant was assigned one camera configuration, but tried all techniques. We selected three locations in our university campus for conducting the experiment, in order to prevent learning effects between repetitions. The three locations appeared different from each other, with different types of buildings and varying density of trees. We used a Latin square to balance the order in which techniques and locations were assigned to each user. Participants' gender and familiarity with the locations were also balanced between camera configurations.

Before the experiment, we collected demographic data, some information on the amount of time users spend with both paper and digital maps, and information on their spatial abilities, for which we used the SBSOD questionnaire [52]. The experiment started with an outdoor introductory session, where the participant got familiar with

[†] With *spatial awareness* we refer to a person's knowledge of her location within the environment, of the surrounding objects, of spatial relationships among objects, and between objects and herself.

[‡] With *mental workload* we refer to a person's demand of attentional and processing resources.

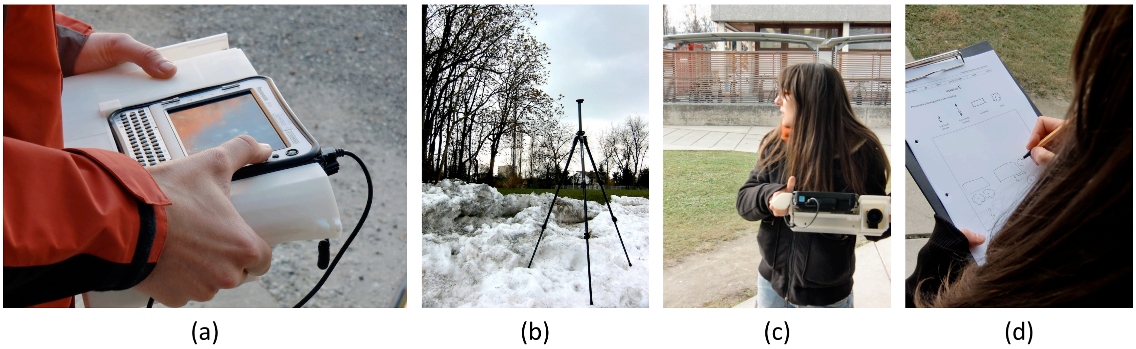


Figure 5.6: The experimental setup. From left to right: (a) the mobile setup used for the experiment, (b) during the experiment, we always positioned one tripod to act as a landmark for the location of the remote camera, (c) one participant searches the remote camera in the environment, (d) throughout the experiment participants were asked to draw maps of the locations.

the handheld device and the techniques. A dummy remote camera was provided for practising.

After the introduction, we blindfolded and walked the participant to one of the three locations. Upon arrival, the participant was provided again with one of the techniques and asked to identify the position and orientation of the remote camera in the environment, by making use of both the device and the physical environment. To prevent user biasing, the techniques were named with neutral names (A, B and C). The participant was allowed to look around but not to move from the designated location. Once the participant felt confident about drawing a map with the main objects in the scene, her position, and the remote camera's location and orientation, we gave her paper and pencil and we took the device. The participant could therefore not use the technique while drawing. Participants also filled in a short questionnaire on their spatial awareness and workload, using parts of the NASA TLX questionnaire [51] and an RSME (Rating Scale for Mental Effort, scale 0–150, 150 for maximum effort) [215]. Finally, we collected a self-reported level of confidence. This procedure was repeated for the other two locations.

After the three techniques were used, we asked the subjects to state their preferred technique for a set of indices related to spatial awareness and workload. In total, participants filled out 12 pages of questions (31 spatial ability questions and 41 ratings based on Likert scale) and drew three maps. The questionnaire is presented in Appendix A.1.

Apparatus. The experimental setup is shown in Figure 5.6. We used a handheld device consisting of an ultra-mobile computer (Panasonic CF-U1 with a 5.6" screen), a Ublox GPS sensor and a uEye UI-2210 colour camera (640 × 480 resolution) mounted with a 4.2MM

Pentax wide-angle lens. The uEye camera was physically bound to the whole setup, and acted as the local camera in the user study. All three techniques were used in combination with pre-recorded video feeds from static remote cameras, to avoid the risk of connectivity problems related to the streaming of the remote video during the experiment. Both the local and remote camera feeds were running at 30 frames per second. Screenshots of the three techniques and the different transition views are shown in Figure 5.3. A tripod was positioned in the field to represent the remote camera. Participants transitioned between the states of a technique through a button on the device.

Participants. A total of 27 users (16 male, 11 female, aged between 22 and 48) participated in the study, so we had 9 participants for each camera configuration. All (but one) participants had normal or corrected-to-normal vision. To partially compensate for the effects of prior knowledge of the environment, we invited 17 users that regularly visited the campus and 10 users that had hardly visited the campus before. Of the 17 regularly visiting users, 9 users hardly knew at least one of the locations. Results were collected from the 27 users and 3 locations, for a total of 81 trials. The duration of the experiment was about 75 minutes per participant.

Results. As a result of the SBSOD questionnaire analysis, we retrieved a median score of 65.56%. Based on the median and the standard deviation of the results, we separated the users in three groups:

G1: low spatial ability (< 55%), 2 male / 7 female.

G2: average spatial ability (55–75%), 9 male / 3 female.

G3: high spatial ability (> 75%), 5 male / 1 female.

Female users were more prevalent in the G1 and G2 spatial ability groups (Pearson correlation, $p < .01$): this is in line with previous studies [90]. It must be noted that the three spatial ability groups are computed using the median and the variance of the spatial ability of all participants. We did not recruit all participants at once, but we rather recruited them while the study was running. Consequently, we could assign each participant to the corresponding spatial ability group only after the experiment was concluded. Therefore, we could not balance spatial ability between the different camera configurations (see Table 5.1), and we cannot make exact statements on spatial-ability effects per camera configuration as no participants with high spatial ability fell within the nCS condition. In the following table, we subdivide the sample population as a function of our three research conditions.

	Gender		Familiar with the locations		Spatial ability		
	Male	Female	Yes	No	G1	G2	G3
CS	5	4	5	4	2	3	4
CnS	6	3	6	3	3	4	2
nCS	5	4	6	3	4	5	0

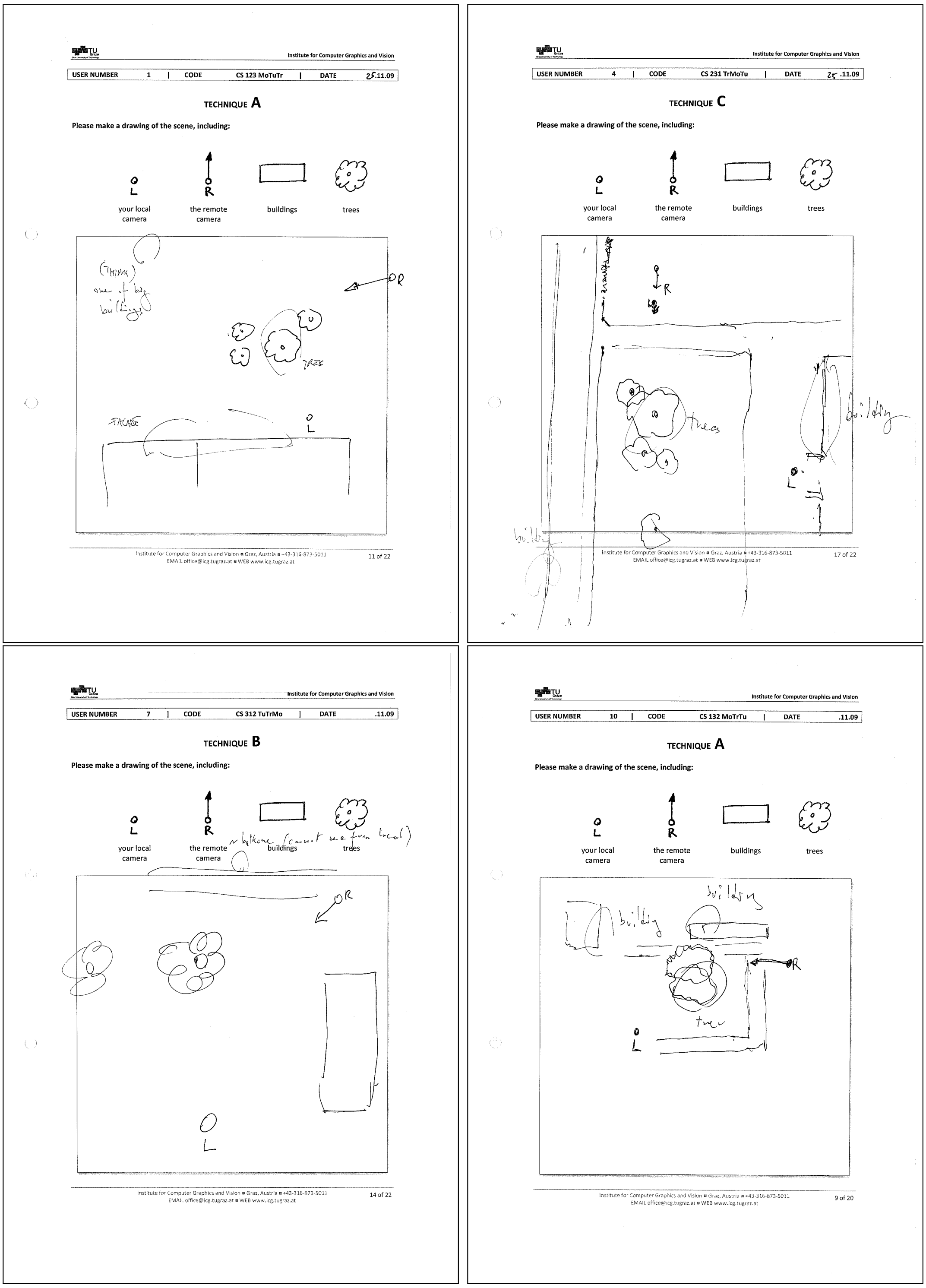
Table 5.1: The sample population for our experiment. Distribution of gender, familiarity with the locations of the experiment and spatial ability among the different camera configurations.

Spatial awareness. We started with interpreting the maps drawn by the users. Due to the diversity and quality of drawing (for an example, see Figure 5.7), we made a qualitative analysis of the errors. We used a voting mechanism among researchers to count errors in the overall spatial configuration of objects in the map (VS), and in the position (VP) and orientation (VO) of the remote camera. The results of this analysis are shown in Table 5.2.

	Mosaic			Tunnel			Transitional		
	VS	VP	VO	VS	VP	VO	VS	VP	VO
CS	0	1	1	3	2	3	2	1	3
CnS	2	1	1	3	1	2	3	1	0
nCS	1	3	0	2	3	3	1	1	1
G1	1	2	0	5	3	3	3	1	1
G2	2	3	2	2	3	3	3	2	1
G3	0	0	0	1	0	2	0	0	2
TOTAL	**10**			**22**			**13**		

Table 5.2: The total number and types of errors in the map drawings for each technique. Results are shown for each camera configuration, each spatial-ability group, and overall.

Mosaic caused the least errors in the drawings. Users made few errors when drawing the remote camera position, even if the technique itself does not provide any distance information. The Transitional technique performed very well in the nCS condition, especially when one considers that no high spatial ability users fell within this condition. The technique seems to provide quite accurate information on the remote camera's placement and orientation when the remote camera is not visible by the user. High-ability participants made significantly fewer errors (Pearson correlation: $p < .01$) than participants from the other spatial-ability groups. Previous knowledge of the environment only had a significant main effect on errors produced by the Tunnel technique, but not on the other techniques



Institute for Computer Graphics and Vision

USER NUMBER	1	|	CODE	CS 123 MoTuTr	|	DATE	25.11.09

TECHNIQUE A

Please make a drawing of the scene, including:

Institute for Computer Graphics and Vision ■ Graz, Austria ■ +43-316-873-5011
EMAIL office@icg.tugraz.at ■ WEB www.icg.tugraz.at
11 of 22

Institute for Computer Graphics and Vision

USER NUMBER	4	|	CODE	CS 231 TrMoTu	|	DATE	25 .11.09

TECHNIQUE C

Please make a drawing of the scene, including:

Institute for Computer Graphics and Vision ■ Graz, Austria ■ +43-316-873-5011
EMAIL office@icg.tugraz.at ■ WEB www.icg.tugraz.at
17 of 22

Institute for Computer Graphics and Vision

USER NUMBER	7	|	CODE	CS 312 TuTrMo	|	DATE	.11.09

TECHNIQUE B

Please make a drawing of the scene, including:

Institute for Computer Graphics and Vision ■ Graz, Austria ■ +43-316-873-5011
EMAIL office@icg.tugraz.at ■ WEB www.icg.tugraz.at
14 of 22

Institute for Computer Graphics and Vision

USER NUMBER	10	|	CODE	CS 132 MoTrTu	|	DATE	.11.09

TECHNIQUE A

Please make a drawing of the scene, including:

Institute for Computer Graphics and Vision ■ Graz, Austria ■ +43-316-873-5011
EMAIL office@icg.tugraz.at ■ WEB www.icg.tugraz.at
9 of 20

Figure 5.7: Some of the maps drawn by our participants. All maps depict the same location and the same camera configuration: there was a great diversity in the drawing style of the different participants.

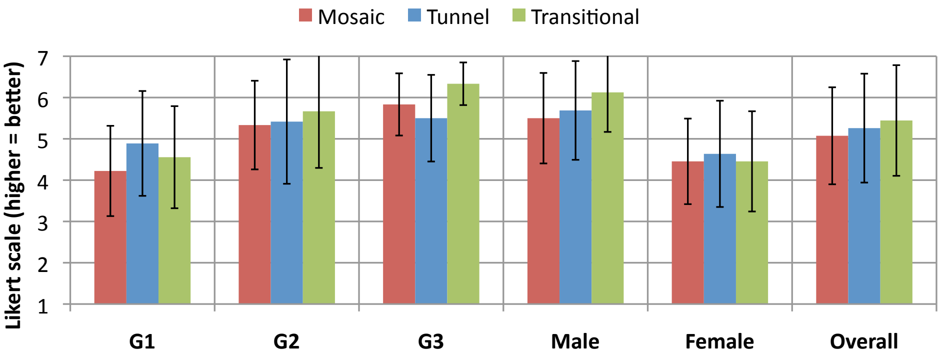


Figure 5.8: Self-assessed success ratings. Scores for each spatial ability group, each gender and overall (error bars = SD), on a 7-point Likert scale (higher scores mean higher self-reported success).

(one-way ANOVA, $F_{1,25} = 9.04$, $p < .01$): users with previous knowledge performed better with the Tunnel technique than users with no previous knowledge.

For each technique, we asked participants to self-assess their success in drawing the map (see Appendix A.1). The results are shown in Figure 5.8. Users with higher spatial ability felt more confident. The correlation between spatial ability and errors is therefore reflected in the subjective assessments. Regarding the techniques, groups G2 and G3 felt most confident with Transitional, whereas group G1 preferred Tunnel. Mosaic caused the least errors, but users did not report the highest confidence in this technique for drawing the map. There was a significant interaction between spatial ability and success rating (one-way ANOVA, $F_{2,24} = 5.17$, $p = .01$). A post-hoc t-test shows a main effect of spatial ability on success rating for Mosaic ($p = .02$) and Transitional ($p = .03$): higher-ability users felt more confident using these two techniques than users with lower ability. Spatial ability did not have a significant effect on the success ratings for Tunnel.

In general, users estimated that with either technique they needed to retrieve as much information from the screen as from the environment itself (see Appendix A.1 for the questionnaire): since the techniques provide only limited information, users also needed to observe the environment to fulfill their task. We expected that Mosaic would require users to observe the environment more than the other interfaces, because this visualisation technique only provides minimal information – the direction of the remote camera. We therefore expected users to integrate this information with information gathered by looking at the environment, in order to identify the physical position of the remote camera. In contrast, we expected that Tunnel would require users to observe the environment less than the other interfaces, because the information that can be gathered from the environment is already integrated in the AR view. However, a repeated measures ANOVA provided no significant difference among techniques ($F_{2,52} = .28$, $p = .76$), suggesting that for all techniques users felt that they paid as much attention to the screen as to the environment.

Overall, Mosaic performs better, producing the least errors but not producing the highest success ratings among users. Transitional gives higher success ratings and a performance comparable to Mosaic (slightly better when the remote camera is hidden from the user). Users with higher spatial ability rate their success higher with either Mosaic or Transitional, while users in the lower spatial ability group rate their success higher with Tunnel, although it produces more errors.

Mental workload. To analyse workload, we considered both the workload-related questions derived from TLX and the RSME scale. We found a direct correlation between TLX and RSME ratings (Pearson correlation, $p < .01$ for all techniques), forming a reliable base to judge the user's workload (Figure 5.9). There is a tendency of Mosaic to require less workload and a tendency of Tunnel to require more, but a repeated measures ANOVA did not show a significant main effect of mental workload on the techniques ($F_{2,52} = .280$, $p = .75$), and a one-way ANOVA did not show an effect between spatial ability and Mosaic ($F_{2,24} = 2.77$, $p = .08$), Tunnel ($F_{2,24} = 1.06$, $p = .36$), or Transitional ($F_{2,24} = 2.57$, $p = .10$). Additionally, no significant effect could be found between group, technique and camera conditions after multivariate analysis. Finally, a repeated measures ANOVA did not show any effects on the order and progress of the test on the mental workload ($F_{2,52} = .27$, $p = .76$). Overall, there is a tendency of Mosaic to require less workload. Although not significant, Transitional also received a better rating than Tunnel.

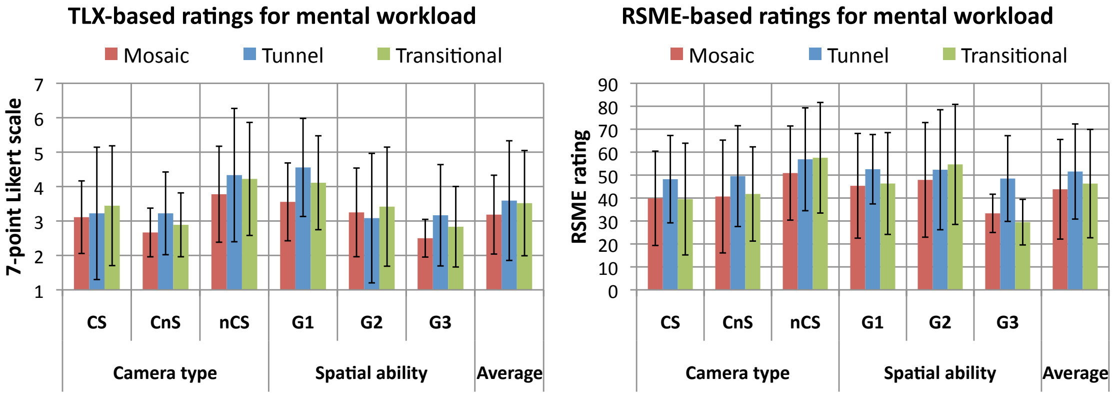


Figure 5.9: Subjective workload ratings. The average and standard deviation of mental load (7 point Likert scale, lower is better) and RSME for each camera type and each spatial ability group, and on average (error bars = +/- SD).

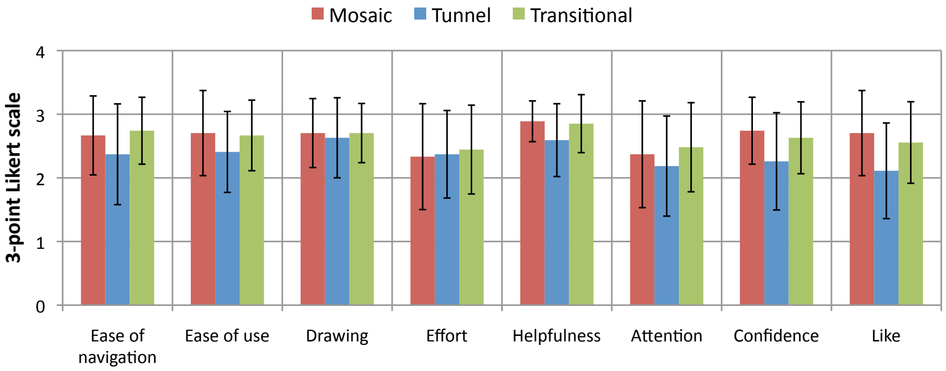


Figure 5.10: The average preference ratings. Scores on a 3-point Likert scale (1–3), where higher is better (error bars = SD).

Subjective preference. Users liked Mosaic best, followed by Transitional (Figure 5.10). A repeated measures ANOVA shows an effect of Technique on user preference: a post-hoc t-test shows that Tunnel was preferred significantly less than Mosaic ($p < .01$) and Transitional ($p = .03$). We noticed no significant effect of spatial ability on the technique preference. Users liked Tunnel less, consistently for all camera conditions. Mosaic was liked most in all camera conditions, whereas Transitional is especially liked in the CnS and nCS conditions. There was no significant differences between the ratings of the techniques for attention, effort, general navigation preferences, as well as the usage of the techniques for drawing a map.

Users gave high ratings to the usefulness of all techniques in helping them to draw a map. A repeated measures ANOVA shows a significant effect in the confidence ratings ($p = .02$): a post-hoc t-test shows that users were significantly more confident in Mosaic than Tunnel ($p = .03$), but there was no significant difference in confidence between the other pairs of techniques. Spatial ability did not have a significant effect on the preference ratings (see Appendix A.1 for the questionnaire). Overall, there is a subjective preference for Mosaic, though it did not always perform significantly better than Transitional.

5.1.3 Discussion

Subjective preferences show that participants found all techniques equally helpful for the task of drawing the maps. However, Mosaic and Transitional generally received higher preference scores, while Tunnel was rated lower. Similarly, participants also felt more confident performing with Mosaic, just a bit less with Transitional, but significantly less confident with Tunnel. Cross-comparing subjective ratings with errors and workload, we can see that Mosaic and Transitional caused less errors (and consequently higher spatial awareness), and imposed slightly less workload. Furthermore, we noticed that the techniques have different robustness to registration errors: Tunnel is less robust to misalignments than the other two techniques, due to the choice of an AR visualisation. When registration errors occurred with Tunnel, participants were forced to observe the environment more closely: this can explain part of the higher workload experienced with Tunnel. Overall, we think that users preferred techniques that let them perform reasonably well while imposing lower workload: in this case, the two non-AR techniques.

5.2 Providing overview on all augmentations

Our previous results highlight the value of combining AR interfaces with compass-like overlays or transitional interfaces, as they are more effective in pointing the user to the information. The previous experiment relies on the assumption that the system knows exactly which piece of information is needed by the user – for example, this is the case when a user formulates a very specific query that has one single result. In other cases, a user might not be able to specify a clear query, or there might be multiple results of interest for the user. In such cases, the interface should provide the user with an overview of all surrounding information. This second part of the chapter extends our previous results to generic search tasks. We consider both the case in which the system can point the user to a specific piece of information, and the case in which the system cannot do so and must provide overview to the user. For this work, we target the scenario of AR browsers – applications that augment the environment with information retrieved from online sources.

5.2.1 Interface design

We propose two zooming interfaces for providing an overview on virtual information (Figure 5.11): an egocentric zoom that increases the field of view up to 360° and an exocentric zoom that smoothly moves between first- and third-person views of the information. Animations support coherence between zoomed-in and zoomed-out views, as suggested by Robertson: *"interactive animation is used to shift some of the user's cognitive load to the human perceptual system"* ([160], p. 190).

Zooming Panorama. The visualisation of the Zooming Panorama is centred on the live video of the phone. As users zoom out, gradually less screen space is used for the live

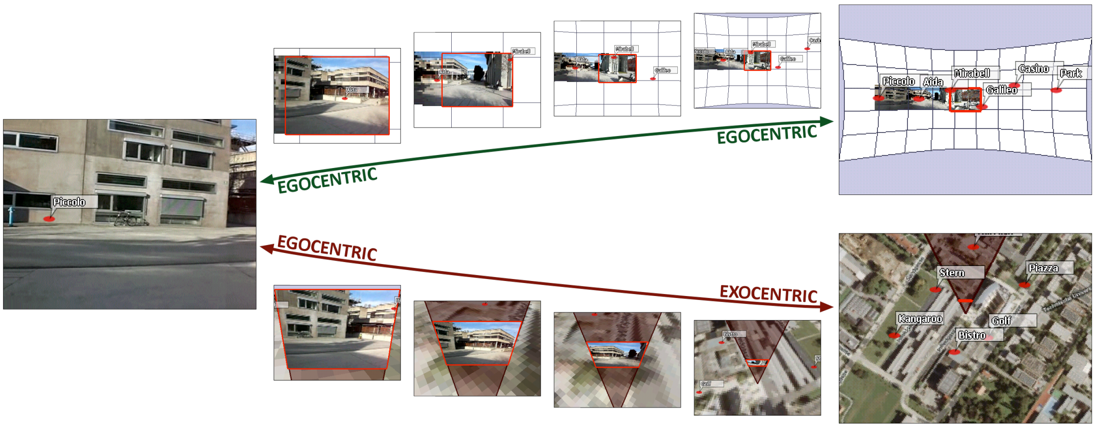


Figure 5.11: The proposed zooming interfaces, as the user turns the camera to the right. We propose an egocentric zoom that increases the field of view up to 360° (top) and an exocentric zoom that gives the user a top-down view onto the information.

video and more space is left for the surrounding panorama (Figure 5.11, top). Information is shown on the screen as text labels indicating the identity of the object. As the camera rotates, we generate a real-time panorama – more information on the tracking technology can be found in Chapter 4. We visualise this panorama as a spatial clue for the user's rotations and as a trace of the visited information. We integrate a wireframe grid to support the understanding of the visualisation as suggested by Zanella et al. [213].

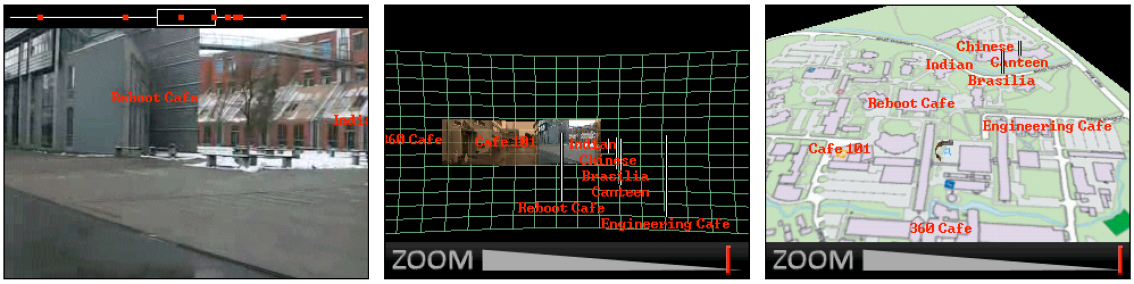
Zooming Map. The visualisation of Zooming Map is always centred on the user's position. As users zoom out, we smoothly animate the camera away from the first-person AR view to an exocentric view presenting a satellite image of the user's surroundings (Figure 5.11, bottom). In the zoomed-out visualisation, we represent the user's position and field of view as a glyph. The satellite image is augmented with text labels corresponding to the information. We use a forward-up map: the user's view direction always corresponds to the top of the screen.

In both interfaces, users can pan the information by physically turning the camera in the environment, just like in AR. Zooming is triggered via an on-screen zoom slider.

Compass. For comparison, we also implemented an overlay-based interface similar to the Context Compass by Lehikoinen et al. [92]. In our implementation, Compass shows the horizontal position of information with respect to the user's view direction around 360°. Similar to the original implementation, Compass is only 1D and the pitch angle of annotations is not represented (all icons lay on the same horizontal line). Due to the similar metaphor and its 2D nature, Compass is comparable to the Mosaic interface evaluated in the previous section of this chapter.

5.2.2 Pilot study

We implemented a first prototype of each of the interfaces to gather preliminary feedback on their usability. Figure 5.12 shows screenshots of the three interfaces in their early form. We used this prototype to conduct a pilot study in the main campus of our university.



(a) (b) (c)

Figure 5.12: The first prototype used for gathering preliminary feedback. (a) Compass, (b) Zooming Panorama and (c) Zooming Map.

Task. We asked participants to explore the cafés in their surroundings using our interfaces.

Independent variable. We had three conditions for this experiment: *Compass*, *Zooming Panorama*, and *Zooming Map*.

Research questions. The study focused on identifying major usability problems of the proposed interfaces, collecting subjective opinions, and comparing the zooming interfaces with Compass.

Procedure. Participants were provided with a one-page information sheet presenting each interface with a screenshot (see Appendix A.2), describing the visualisation and the interaction. The prototype was programmed with geo-referenced positions of all cafés and restaurants on campus.

After reading the information sheet, participants were asked to spend some minutes with each interface separately. As shown in Appendix A.2, participants were not given a specific task to complete, but they were asked to explore the locations of all the cafés and restaurants around them in the campus, with respect to their location on the campus. We balanced the order in which the interfaces were shown to different participants. Our main interest was letting participants experience the potential advantages and shortcomings of each interface.

After the experiment, we conducted a semi-structured interview with each participant, asking for opinions about pros and cons of each interface and pair-wise comparisons between the interfaces. During the interview, all participants were again provided with the information sheet containing the screenshots of the user interface, in order to support their memories.

Apparatus. The prototype runs at interactive frame rates (approximately 10 frames per second) on an HP iPAQ 614c phone (on Windows Mobile). Due to the lack of an accelerometer and a compass in the mobile phone, the tracking technology of our prototype relied solely on optical flow from natural features [199]. Zooming was implemented using a slider on the bottom of the screen, as shown in Figure 5.12.

Participants. Since we required users to actively help us spotting initial usability issues, we recruited five participants with experience in user interface design and augmented reality. The participants had background in computer science (2), architecture (1), psychology (1) and arts (1).

Results. In the following, we separately discuss the feedback gathered for each interface.

Zooming Panorama. The panorama trace left by the moving camera (Figure 5.12 (b)) was generally found to be a useful cue: as users turned the camera in the environment, we automatically captured a panorama and visualised it as a cue for the device's movement in the information space. However, users commented that the panorama is "too dim" and merged with the black background, in particular under outdoor lighting conditions. Users were also concerned about the screen space occupied by the context in the zoomed-out view, claiming that it is hard to see the details in the video and in the panorama trace.

A recurring comment on the Zooming Panorama is that it encodes only direction information, while it does not provide information about distance and occlusions. Two users suggested considering adding representations for distance and occlusions (for example through colour or transparency), commenting that a textual representation of the distance might clutter the view.

Users pointed out similarities between Zooming Panorama and Compass. Comparing the two, users noted both advantages and disadvantages in our proposed technique. While Compass is limited in telling you where off-screen objects are, Zooming Panorama also conveys what these objects are. Zooming Panorama takes more screen space and therefore the details can be hard to read when zooming out. Two users suggested combining the strengths of the two interfaces by showing Compass when Zooming Panorama is completely zoomed in, in order to still provide some coarse context information. Comparing Compass against the other two interfaces, one user pointed out its advantage of not requiring user input and being always visible on screen.

Zooming Map. Most users stressed the importance of the transition being smoother. One user suggested exploiting the smooth transition to "make the user aware that [the exocentric view] is a map, and it is flat or tilted down". A strong advantage of this interface seems to be its capability to show "the actual location, not just the orientation" of places, with words such as "distance" and "depth" spontaneously recurring in most interviews. Two users commented that Zooming Map would be most useful when "you're stuck in a city [...] not knowing what's around you" or "in a really difficult navigation space".

There were two contrasting opinions on Zooming Map. One user said that she found the system easy to use because it fits the real-world experience that users already have with maps: in particular, this was put in contrast to Zooming Panorama that adopts a non-familiar metaphor (a distorted panorama). A further advantage was seen in the fact that the map is user-centred and self-orienting (forward-up) making it quite different from a static map. A second user, however, was skeptical about the use of a map, saying that the technique might not be suitable for people that have problems with using real maps.

Zooming Map was the most appreciated by all users. When comparing it with the other two interfaces, users generally claimed that their preference for it depends on the better level of overview provided by the map, in particular when a desired point of interest is not directly visible from the user's position. The ability to convey a measure of the distance from the user was also considered significant. Interestingly, the study revealed that most of the time users only exploited completely zoomed in and completely zoomed out views.

5.2.3 Refined interface design

We refined our prototype using the feedback from the pilot study (Figure 5.13). We modified Zooming Panorama in order to increase the contrast: a new bright background, with darker grid lines, improves the prominence of the panorama trace also under outdoor lighting conditions. The details in the panorama are still too small, but we did not investigate possible solutions to the issue yet: we exploit the panorama as a spatial trace rather than an augmentation target. As suggested by our users, we integrated Zooming Panorama and Compass into a single interface. We slightly modified the compass to also provide vertical displacement of information, since we noticed that the previous 1D representation required users to blindly search the information on the vertical axis. We changed the background to a bright colour for the Compass too. Figure 5.13 (a–b) shows the new implementations, and how Zooming Panorama and Compass are combined and synchronised: the central rectangle that represents the current field of view of the user in Compass always matches the field of view of Zooming Panorama. We did not modify Zooming Map but rather optimised it in order to speed up the rendering on phone, making the animation smoother.

Figure 5.14 shows how users can interact with the new interface. Zooming the

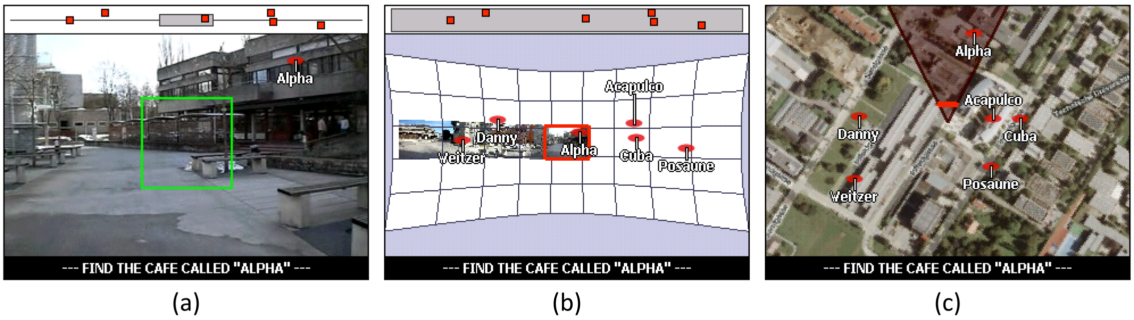

Alpha
--- FIND THE CAFE CALLED "ALPHA" ---
Acapulco
Danny
Weitzer
Alpha
Cuba
Posaune
--- FIND THE CAFE CALLED "ALPHA" ---
Alpha
Acapulco
Danny
Cuba
Posaune
Weitzer
--- FIND THE CAFE CALLED "ALPHA" ---
(a)
(b)
(c)

Figure 5.13: The second prototype of our interfaces, used for the two controlled studies. (a) Compass, (b) Compass and Zooming Panorama, and (c) Zooming Map. In the controlled studies, the user's task was always presented on the bottom of the screen (the green viewfinder in (a) represents the area for selecting the target object in order to complete the task).

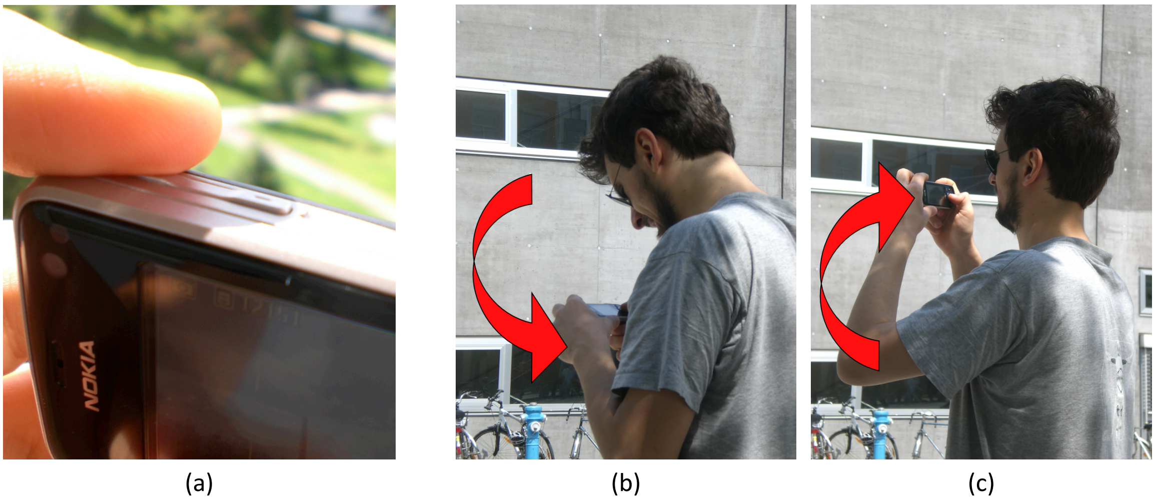


Figure 5.14: Zooming in our second prototype. (a) Zooming Panorama is triggered using the zoom button of the camera phone; Zooming Map is triggered by tilting the phone down (b) for the exocentric view and up (c) for the egocentric view.

panorama is triggered via the zoom buttons of the phone – this is a more intuitive metaphor than the on-screen slider, and is similar to the zoom in a regular digital camera. Based on the observation that users mostly use the completely zoomed-out or zoomed-in views of Zooming Map, rather than intermediate viewpoints, we adopted a gesture-based approach: tilting the phone down zooms to the exocentric map view while tilting it up zooms back to full-screen AR.

5.2.4 Controlled study 1

Using the refined prototype, we conducted a study to investigate how our zooming interfaces perform in search tasks representative for the AR-browser scenario. We compared our interfaces against Compass, designing the study to answer the following research question: *do zooming interfaces allow shorter task completion times and smaller travelled distances than a Compass interface?* With *travelled distance*, we refer to the amount of turning of the device's camera in space to access information through AR.

Tasks. We asked participants to perform two different search tasks:

T1. Find a well-defined café, with a name known a priori.

T2. Find the closest café, thus considering all information before making a choice.

Independent variables. The study had one four-way and two two-way independent variables:

- *Interface.* We designed four interfaces, to cover different combinations of the three degrees of freedom of information surrounding a user – yaw, pitch and distance. We chose the four interfaces as Compass (yaw), Compass + Zooming Panorama (yaw, pitch), Compass + Zooming Map (yaw, distance), Compass + Zooming Panorama + Zooming Map (yaw, pitch, distance). We will refer to them respectively as **C**, **CP**, **CM** and **CPM**.
- *Task.* The two tasks **T1** and **T2** described above.
- *Density.* We chose two levels of information density: either **6** or **12** cafés.

Hypotheses. We formulated the following hypotheses:

H1. The zooming interfaces allow faster task completion times. We expected that the higher amount of overview offered by the zooming interfaces would allow users to find the target café just by looking at the screen. In contrast, we expected Compass to require participants turning the device towards each annotation, until they hit the correct one.

H2. The zooming interfaces require travelling smaller distances compared to Compass. Similar to H1, we expected that users would move less since the zooming interfaces provide a more informative overview, and all information is accessible in the zoomed-out view without the need of turning the device's camera in space.

H3. No difference in completion time and distance between 6 and 12 cafés for the zooming interfaces will be evident, while Compass will require more time and longer travelled distances with 12 cafés than with 6 cafés. Since completing the task with Compass requires pointing the device towards all annotations, one by one, we expected that task completion time with Compass would increase with the growing number of annotations.

Procedure. For each Interface and Task, participants performed one practice trial and then two repetitions per information Density. This resulted in 4 (Interface) × 2 (Task) × 2 (Density) × 2 (Repetition) = 32 observations per participant. Every participant used all four interfaces, but the order followed a balanced Latin square to reduce carry-over effects. A complete experimental session took approximately 45 minutes.

We represented the location of cafés as text labels containing the name and, for task T2, the distance from the user. To prevent learning effects and prior knowledge, we randomised café names and locations at each repetition, with distances varying between 40 and 160 meters.

Participants started each repetition by pressing the phone's central button, turning the phone in the direction of the correct café and pressed the central button again to complete the repetition (wrong presses were ignored). All users were instructed to work as fast and as accurate as possible. We logged all task completion times and the rotations of the camera (as angular distances), considering the latter as travelled distance in the information space.

Apparatus. The refined prototype runs at approximately 15 frames per second on a Nokia 6210 Navigator phone. We fuse sensor- and vision-based tracking to estimate the orientation of the device (see Section 4.2) to obtain more accurate measurements compared to the phone's sensors alone. The phone's accelerometer is also used to detect the tilts that trigger Zooming Map.

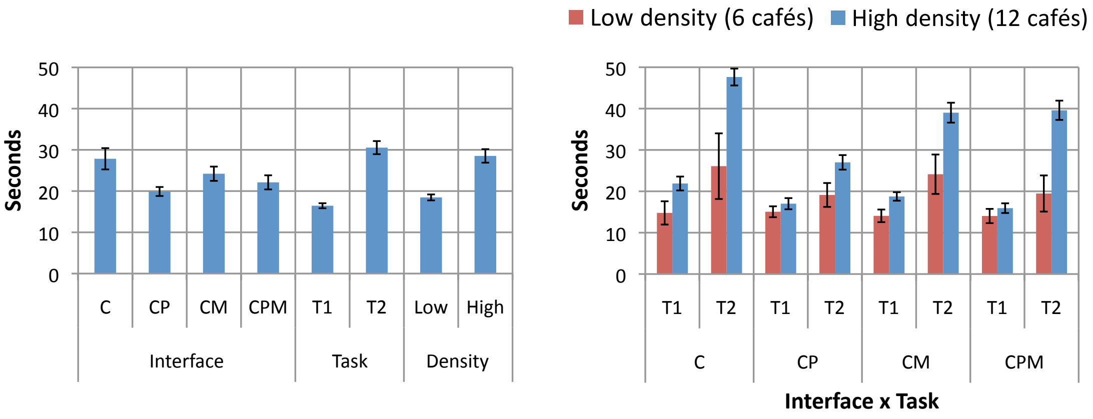


Figure 5.15: Task completion times (in seconds) for the first user study. Left: average task completion time. Right: task completion time for each Interface, Task and Density (error bars = +/- SE).

Participants. Twenty university students (10 female and 10 male) aged between 23 and 34 (Mean (M) = 27.35, SD = 3.10) participated in the study. All participants had normal or corrected-to-normal vision.

Results. The collected data shows some outliers (5% for task completion time, 3.75% for distance). During the study, we observed several instances in which the label-positioning algorithm failed, causing inconsistencies between subsequent frames. Some users commented on the issue at the end of the study. We observed that this slowed down participants when it occurred, and we think that most of the outlying points in the collected data are attributable to this problem. For all extreme outliers (3 × inter-quartile range), we kept the single non-outlying measurement of the two repetitions rather than their average, and dropped the outlying measurement.

For each task repetition, we calculated the distance ratio d_u as the ratio

$$d_u = \frac{d_m}{d_0} \tag{5.1}$$

between d_m, the distance effectively travelled by the user (in radians), and d_0, the shortest distance between the user's initial orientation and the direction of the target (also in radians). The measure that we obtain is therefore a multiple of the shortest distance, where the ideal value is 1 (meaning that the user travelled the shortest possible distance).

We analysed the effects on time and distance ratio with a 4 (Interface) × 2 (Task) × 2 (Density) repeated-measures ANOVA.

Task completion time. All main effects are significant, as well as the interactions Interface × Task and Task × Density (Figure 5.15). Comparing the interfaces ($F_{2,36.07} = 4.66$, $p = .02$) we found longer task completion times for C ($M = 27.84$, $SE = 2.58$) compared to CP ($M = 19.90$, $SE = 1.09$), while CM ($M = 24.21$, $SE = 1.73$) and CPM ($M = 22.12$, $SE = 1.71$) did not differ significantly from the other interfaces. Task T2 took longer on average ($M = 30.54$, $SE = 1.59$) than task T1 ($M = 16.45$, $SE = 0.61$) ($F_{1,18} = 44.67$, $p < .01$) and high density ($M = 28.53$, $SE = 1.64$) took longer than low density ($M = 18.47$, $SE = 0.72$) ($F_{1,18} = 40.78$, $p < .01$).

The interaction Interface × Task ($F_{3,54} = 2.79$, $p = .05$) showed a significant main effect. A post-hoc t-test showed no difference between the interfaces for task T1, but for task T2, CP was faster than C ($t(19) = 3.59$, $p = .002$) and CM ($t(19) = -2.90$, $p = .009$). For all interfaces task T2 took longer than task T1. The Task × Density interaction ($F_{3,18} = 17.33$, $p < .01$) showed a relatively steep increase in completion time for task T2 ($t(19) = -5.97$, $p < .001$) as the information density increases (+73%, 22.2 seconds for low density and 38.3 seconds for high density) compared to a smaller increase for task T1 ($t(19) = -3.87$, $p = .001$) (+27%, 14.5 seconds for low and 18.4 seconds for high density).

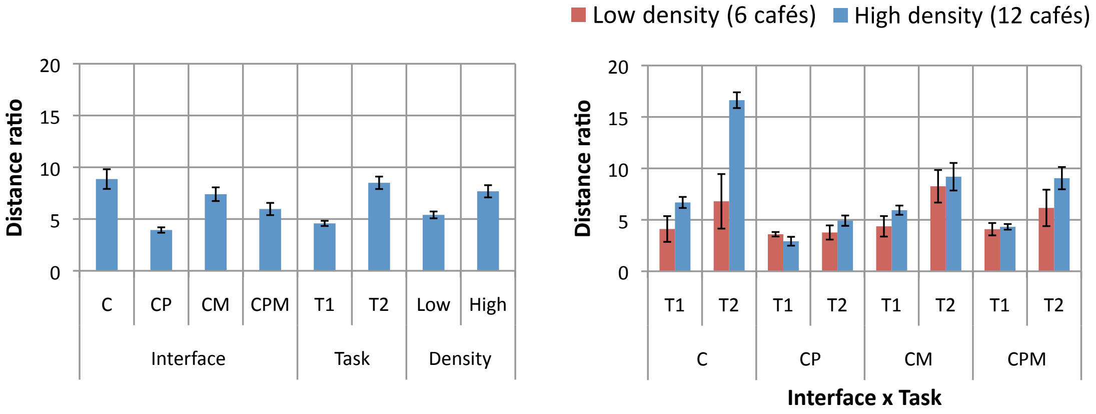


Figure 5.16: The distance travelled for the first user study. A distance ratio is computed as a multiple of the shortest distance (see Equation 5.1). Left: average distance travelled. Right: distance travelled for each Interface, Task and Density (error bars = +/- SE).

Distance ratio. All main effects were significant, as well as all two-way interactions (Figure 5.16). Comparing the interfaces ($F_{3,54} = 10.62$, $p < .01$), we found that with CP participants moved less ($M = 3.94$, $SE = 0.26$) than with C ($M = 8.86$, $SE = 0.95$), CM ($M = 7.40$, $SE = 0.66$) and CPM ($M = 5.97$, $SE = 0.59$). Task T2 showed almost double distance ratio ($M = 8.50$, $SE = 0.60$) compared to task T1 ($M = 4.58$, $SE = 0.25$)

($F_{1,18} = 46.18$, $p < .01$). Participants also moved the phone more with 12 cafés ($M = 7.68$, $SE = 0.59$) than with 6 cafés ($M = 5.41$, $SE = 0.33$) ($F_{1,18} = 18.34$, $p < .01$).

The interaction Interface × Task ($F_{3,54} = 2.77$, $p = .05$) showed a significant main effect. A post-hoc t-test showed no difference between the interfaces for task T1. For task T2, however, we found that CP required smaller distance ratios than C ($t(19) = 3.89$, $p = .001$) and CM ($t(19) = -3.20$, $p = .005$). CP was the only interface not showing a difference between the two tasks, while all other interfaces showed higher distance ratios for task T2 than for task T1. For the interaction Interface × Density ($F_{3,54} = 5.86$, $p < .01$), we found smaller distance ratios in the low-density case for CP compared to C and CM, while for high density CP was only different from C. The interaction Task × Density ($F_{3,18} = 11.68$, $p < .01$) showed a bigger increase in distance ratio between density levels for task T2 (+60%, 6.2 for low and 9.9 for high density) than for task T1 (+25%, 4.0 for low density and 5.0 for high density).

Questionnaire. After using each Interface, participants filled in a questionnaire (see Appendix A.3) containing questions on ease of use, usefulness, and amount of information on the screen. Each statement was answered on a 6-point Likert scale ranging from “completely disagree” to “completely agree”. The questions and the results are shown in Figure 5.17. All significant main effects and pair-wise differences are shown in Table 5.3.

Participants gave relatively high scores for ease of use of all tested interfaces, for the low-density conditions. For the high-density conditions ratings were lower, with the CM interface getting the lowest rating. In both cases, however, the differences were not significant. In terms of usefulness to complete the task, C received lower scores compared to the other interfaces in both the low and the high-density case. For the low-density case, there was a significant difference between C and CP, and between C and CPM; the difference between C and CM is only marginally significant. For the high-density case, there is a significant difference only between C and CP, while the difference between C and CPM is only marginally significant. The two questions regarding the amount of information on the screen show more distinct patterns. For low-density tasks (6 cafés), participants rated that there was neither too much nor too little information on the screen for all interfaces but C, where there was a slight trend towards too little information (the differences between C and all other conditions, on the question regarding too little information, were all significant). For high density, the Compass shows slightly too little information, while the interfaces containing the Zooming Map tend to have slightly too much information (the differences between C and all other conditions are significant for both questions).

		Main effect (Friedman)	Post-hoc (Wilcoxon with Bonferroni correction, α = .008)					
			C–CP	C–CM	C–CPM	CP–CM	CP–CPM	CM–CPM
Easy to use	6	$\chi^2(3)$ = 7.34 p = .06	-	-	-	-	-	-
	12	$\chi^2(3)$ = 6.47 p = .09	-	-	-	-	-	-
Useful	6	$\chi^2(3)$ = 20.14 p < .01	p = .001	p = .009	p = .002	p = .163	p = .480	p = .046
	12	$\chi^2(3)$ = 13.84 p < .01	p = .002	p = .042	p = .010	p = .015	p = .512	p = .134
Too much info	6	$\chi^2(3)$ = 2.33 p = .50	-	-	-	-	-	-
	12	$\chi^2(3)$ = 20.19 p < .01	p = .003	p = .001	p = .002	p = .270	p = .342	p = .857
Too little info	6	$\chi^2(3)$ = 20.01 p < .01	p = .003	p = .002	p = .001	p = .904	p = .340	p = .518
	12	$\chi^2(3)$ = 25.82 p < .01	p = .001	p = .002	p = .001	p = .209	p = .516	p = .233
Animation		$\chi^2(2)$ = .929 p = .63	-	-	-	-	-	-

Table 5.3: The significant differences in the questionnaire scores. For each question, we report the significance level of the main effect and, if this was significant, the significance levels of all pair-wise differences between conditions. We colour code the significant differences (green), the marginally significant ones (yellow), and the non significant ones (red).

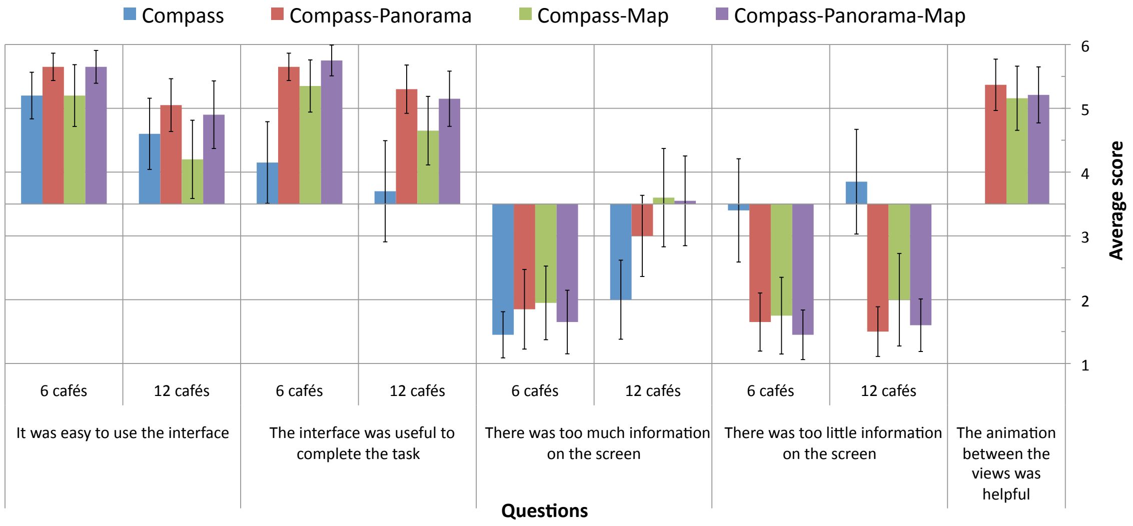


Figure 5.17: Questionnaire results for the first user study. The subjective user ratings were given on a 6-point Likert scale, ranging from 1 (strongly disagree) to 6 (strongly agree). Error bars represent 95% confidence intervals.

Participants indicated that the animation between the views was generally helpful.

Finally, we asked participants to rank the four interfaces from most preferred to least preferred. CPM was ranked first (M = 1.4, on a range from 1 (best) to 4 (worst)), followed by CP (M = 2.2), CM (M = 2.7), and last C (M = 3.8). A Friedman test shows that the effect is significant ($\chi^2 = 35.22$, df = 3, $p < .01$). Post-hoc Wilcoxon tests with Bonferroni correction ($\alpha = .008$) show that all differences are significant ($p < .008$), apart from the difference between CP and CM (Z = -1.60, p = .11) and the difference between CP and CPM (Z = -2.01, p = .04).

We asked participants for open comments on their rankings of the interfaces, and any other feedback. Despite the low scores for Compass, participants find this interface good for guidance and confirmation, in particular to keep track of annotations when zooming in from the Zooming Panorama. One further advantage of Compass is that it is less cluttered than the other interfaces, although also less informative. In contrast to the Zooming Map, Zooming Panorama is found more helpful for finding the pitch angle of the annotations, which could be a clear advantage in real-world scenarios. Only one participant reported fatigue after the experiment.

Discussion. Overall, the Compass condition was ranked significantly worse than all other conditions, and the questionnaire scores show that people feel it provided less information than the other interfaces. However, quantitative results are more vague and do not directly support these considerations. In particular, we noticed that Compass was only significantly worse than the CP condition for task 2, and not significantly worse than CM or CPM. Further, for task 1 there was no significant difference between conditions.

The interaction Interface × Task suggests that a difference in performance between the interfaces strongly depends on the specific task. The zooming interfaces might be advantageous only for complex search tasks in which people need overview of all information. In contrast, the Compass might be advantageous for simpler search tasks. This consideration is backed up by informal feedback received from our participants, on the possible benefits of Compass in the case of a café with a known name. In particular, participants asked us why we did not highlight the correct answer in Compass, since the system knew which one it was. Participants believed that highlighting the correct answer would have strongly improved their performance with Compass. We reflected these considerations in a following experiment, in which we looked more closely at the comparative performance with the different interfaces for various types of search tasks.

5.2.5 Controlled study 2

In light of the positive performance of a compass-like overlay in Section 5.1, where the task of the interface was to point users to a specific piece of information, we concluded that Compass might be more suitable for pointing to single pieces of information, whereas the zooming interfaces might be better for tasks requiring information overview. This consideration is also supported by the interaction Interface × Task found in the previous experiment, which suggests that the difference in performance between the interfaces is dependent on the specific task. We therefore designed a second experiment to investigate task-dependent differences in completion time and travelled distance between each separate interface.

Task. We designed two search tasks:

T3. Find a highlighted object.

T4. Find the closest object.

Hypothesis. We formulated the following hypotheses:

H4. The zooming interfaces are slower than Compass if the object is highlighted, but they require equal amounts of distance travelled. This is because the zooming interfaces introduce a latency due to the zooming animation, while Compass has no latency. Further, in this case Compass should provide sufficient information for quickly finding the correct object, because the task does not require overview but only knowledge of the direction of the highlighted object.

H5. The zooming interfaces are faster and require less distance travelled than Compass in task T4, when an overview of the data is necessary. This is because, like in study 1, the overview offered by the zooming interfaces should allow users to find the correct object by just looking at the screen. In contrast, Compass would require users to turn the device towards each annotation, one by one, until they hit the correct one.

Independent variables. One three-way and two two-way independent variables:

- *Interface.* We designed the interfaces as Compass (**C**), Zooming Panorama (**P**) and Zooming Map (**M**), to evaluate the differences between each separate interface.

- *Task.* The two tasks **T3** and **T4** described above.
- *Density.* We again included two levels of information density, with either **6** of **12** labels.

Procedure. The procedure was the same as in user study 1. For task T4, the distance was again represented as a text label.

Apparatus. The apparatus was the same as the one in user study 1. In the prototype used for this study, we fixed the labeling issue that occurred in the previous experiment.

Participants. Ten people (5 male / 5 female) who participated in the first user study were asked to join this second study.

Results. The effects of the experimental conditions on task completion time and distance ratio were analysed with a 4 (Interface) × 2 (Task) × 2 (Density) repeated-measures ANOVA. The distance ratios were again calculated as multiples of the shortest distance to the target label (see Equation 5.1). Figure 5.18 presents the average completion time (left) and the distance ratios (right) for each combination of Task and Interface.

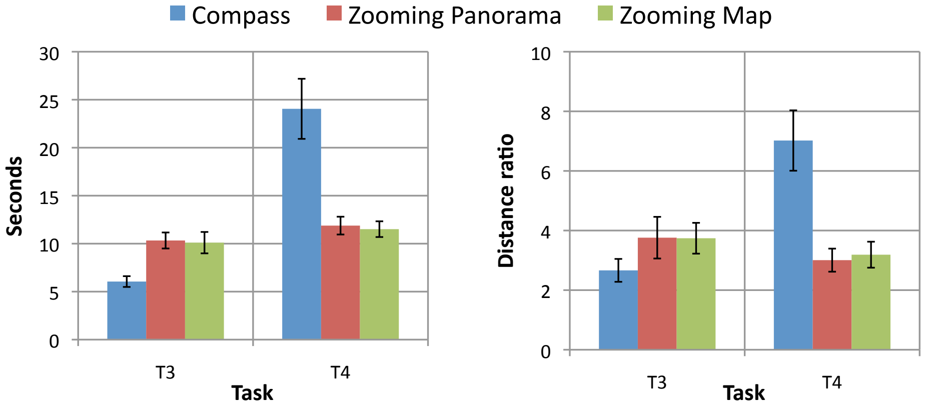


Figure 5.18: The time and distance travelled in the second user study. We show the task completion times (left) in seconds and the distance ratios (right) as a multiple of the shortest possible distance (see Equation 5.1), for each Task and Interface (error bars = +/- SE).

Task completion time. We found significant main effects for Interface and Task, and a significant interaction between the two. Task T4 took on average longer (M = 15.81, SE = 1.24) than task T3 (M = 8.83, SE = 0.49) ($F_{1,9} = 45.86$, $p < .01$). The main effect of Interface ($F_{2,18} = 8.57$, $p < .01$) showed that overall C (M = 15.05, SE = 2.00) took longer than M (M = 10.81, SE = 0.67). P did not differ from the other interfaces (M = 11.11, SE = 0.50). A closer look at the interaction Interface × Task ($F_{2,18} = 25.47$, $p < .01$) revealed that C was faster than the other two interfaces for task T3 and it was slower for task T4. Comparing performance between the two tasks for each interface shows a significant difference in completion time between tasks for C, but not for the two zooming interfaces.

Distance ratio. We found a significant main effect for Interface ($F_{2,18} = 13.65$, $p < .01$) and a significant interaction between Interface and Task ($F_{2,18} = 19.37$, $p < .01$). Post-hoc analysis for the main effect showed overall a longer distance travelled for C compared to the other two interfaces. A closer look at the interaction Interface × Task showed that in task T3 participants travelled smaller distances with C than with M. For task T4, the distance travelled with C was higher than with the other two interfaces. Comparing performance between the two tasks for each interface shows that there was only a difference between tasks for C, but not for the two zooming interfaces.

Observations. Finally, during the two studies we observed a few occurrences in which the tilting movement used for triggering Zooming Map conflicted with other actions that users normally perform. If some information is rather low on the vertical axis (for example, a café in the basement) users must tilt the phone down to see such information, thus triggering an undesired Zooming Map transition. We also observed that some users tilted the phone down when looking at the Zooming Panorama visualisation, probably because of outdoor lighting conditions, since the tilted-down position helped blocking reflections on the screen. This highlights some limitations of our gesture-based approach: for tasks where the phone must be tilted down such interaction would not be an appropriate solution.

5.2.6 Discussion

While we initially hypothesised that users would be faster with the zooming interfaces than with the Compass (H1), our study results only partially support this. In the first study we found the Compass to be slower than the combination of Compass and Zooming Panorama, but users where not significantly faster when Zooming Map was available. In the second study, Compass took twice as long as Zooming Map, but did not differ from Zooming Panorama. We also initially expected Compass to require longer travelled distances than the zooming interfaces (H2). We indeed found that Compass required significantly more movement than Zooming Panorama, but there was no difference between the other interfaces. Finally, we assumed we would observe a smaller increase in time and distance travelled with the zooming interfaces than with Compass, if information density increases (H3). We did not expect a difference between the two zooming interfaces. In the first study we found that high-density tasks took longer in general, independent of the interface used. In the second study we could not observe an effect of information density. Thus our data does not support this hypothesis.

We found the most revealing results in the interaction between Interface and Task: this suggests that the effectiveness of each interface depends on the specific task the user is engaged in. In the second study we hypothesised that the zooming interfaces would be slower than Compass to guide the user in pointing the camera towards a highlighted target object, because of the latency involved in zooming, but we did not expect any differences in distance travelled (H4). For a task requiring users to gain an overview of all information, we expected the zooming interfaces to be faster and to require less travelled distances than Compass, since they provide a more informative overview at a glance (H5). Both hypotheses were supported in our study, apart from Zooming Map, which required more travelling than Compass for a highlighted object. It should be noted, however, that Zooming Map requires users to tilt the phone down to see the map, an additional movement which sums up to the total distance travelled.

Overall, Compass was faster than the zooming interfaces with highlighted objects, but slower in a more complex search task. With Compass, participants took almost four times longer for task T4 than for task T3, whereas the increase in time between the two tasks was only 15% for Zooming Panorama and 14% for Zooming Map. This shows how user performance with the zooming interfaces was less sensitive to the increase in task complexity.

5.3 A digression on the design space of panoramic overviews

In the previous experiments, we showed that zooming between AR and a panoramic overview is beneficial for complex search tasks that require information overview. However, the design of Zooming Panorama visualises the panorama as an almost-rectangular shape. This shape is a common choice for visualising panoramas as well as any other photograph. If we use the term *readability* qualitatively, referring to the ease of reading the visual elements in the panoramic image, rectangles offer a very good readability of the panorama. However, they do not clearly represent the fact that the environment surrounds the user. In particular, the leftmost and rightmost points in the Zooming Panorama depict a part of the environment that is located behind the user. In our AR-browser scenario, users try to understand where augmentations on the panorama are physically located in the surrounding environment. The spatial mapping between panorama and environment must be therefore clear, and our previous experiments gave no clear indication of whether users understood the spatial mapping between the panorama and environment from the rectangular shape.

In this section, we make a digression from AR and investigate the design space of panoramic overviews. In particular, we look at alternative shapes that better communicate the spatial mapping between the panorama and the surrounding environment. The outcomes of this research clearly map back to our work on the Zooming Panorama interface, highlighting how its design can be improved. However, the general impact of this research goes beyond AR and touches all location-based services that use panoramic imagery to support navigation, such as Google Maps Mobile [44].

5.3.1 Interface design

The choice of a shape for a panoramic overview is guided by two design factors: *good readability* of the panorama, and *clear representation of the spatial mapping between the panorama and the environment.* For our experiments, we consider three shapes that make different compromises on these two factors (Figure 5.19).

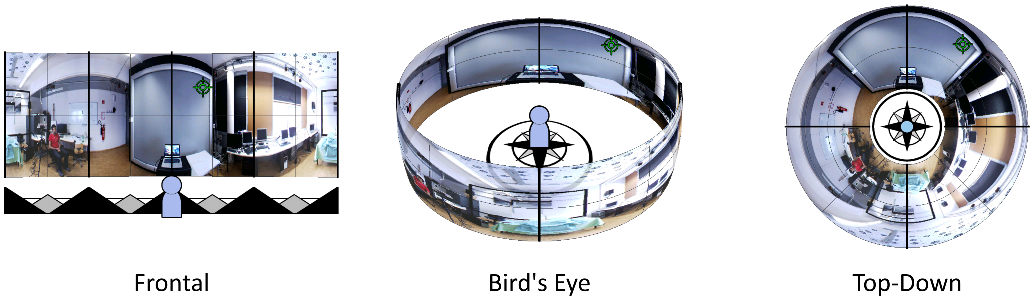


Figure 5.19: 360° panoramic overviews of a laboratory room. We consider three different shapes for visualizing the panorama.

Frontal. The rectangular shape provides good readability of the panorama, but the mapping to the environment is not direct and the visualisation has a discontinuity at the sides.

Bird's Eye. The cylindrical shape maps panorama and environment in a direct way, but readability is sacrificed: due to the 3D view, the panorama appears warped and partially occluded on the sides (we use 15% transparency to resolve occlusions).

Top-Down. A circular shape provides a compromise between readability of the panorama and direct mapping to the environment. Since the pitch is mapped to the distance from the centre of the circle, the panorama appears distorted for high pitch and upside-down in the area of the panorama which is behind the user.

As shown in Figure 5.19, we further enhance all shapes with extra cues, visualising an avatar to represent the position of the user with respect to the panorama, a compass-shaped icon (a *wind rose*) to indicate left, right, front and back directions, and a grid to communicate the distortion of the panorama in the visualisation.

5.3.2 Pilot study

We conducted a pilot study to gain a first understanding on how users exploit panoramic overviews to locate points in the environment. We tested four conditions, defined by the two two-way variables *Shape* (Bird's Eye or Frontal) and *Image* (without a panorama, as in Figure 5.20, and with a panorama of the room, generated using a Ladybug camera[§], as in Figure 5.19). For the experiments, we used a panorama with a resolution of 2048 × 512 pixels.

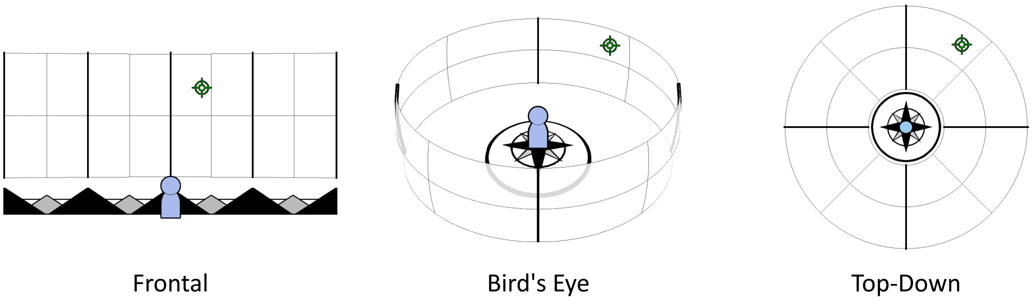


Figure 5.20: The three shapes in the condition with no panorama.

We recruited 4 participants from our university. We showed them a sequence of panoramas, on which we marked one point with a green crosshair, and asked them to point to the corresponding location in the environment. All participants used all conditions. We counterbalanced the order of conditions with a balanced Latin square and randomised the crosshair position.

To isolate and accurately measure comparative performance with the visualisations, we chose a controlled lab setup. We showed panoramas on a 13.3" screen, in a window of 1024 × 768 pixels. Participants pointed with a wand and then clicked a button on the

[§] Ladybug is a product of Point Grey: `http://www.ptgrey.com/`

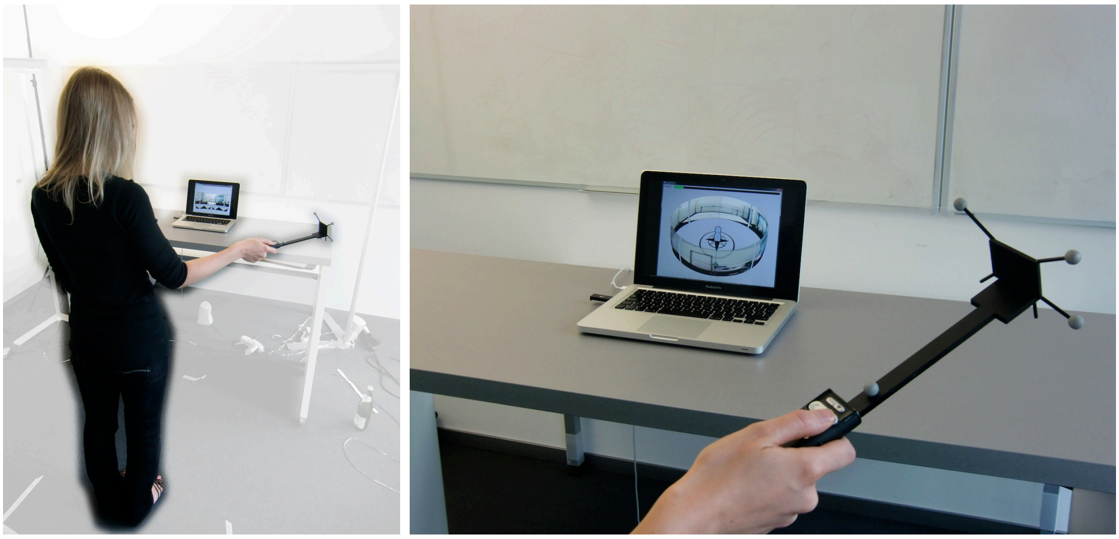


Figure 5.21: The experimental setup. The setup for our user studies consisted of a static display and a tracked wand (used for pointing in the environment).

wand when they thought they were pointing in the right direction (clicks were recorded by a laptop via a wireless receiver). We used a wand to acquire accurate motion and timing data via an infrared tracker[¶]. The setup is shown in Figure 5.21.

After the experiment, we interviewed participants and asked them to describe what strategies they used to locate the points with each different shape (see Appendix A.4). In the interviews, participants described two different types of strategy. The first strategy was to *determine in which body-aligned direction* (left, right, front, back) the point was located; participants reported using mainly the grid and the wind rose for this strategy. The second strategy was performing *visual matching* of objects in the panorama with corresponding objects in the real environment, in order to precisely locate the given point in function of the objects in the environment.

We used the feedback from the pilot study to frame our subsequent user studies. We decided to conduct three separate experiments, in order to isolate the effects of the two strategies: we conducted a first study with no panorama (no visual matching), while the following two studies where conducted with a panorama (to support visual matching). During the three experiments, we collected questionnaire data on expertise with panoramas (browsing, capturing), maps and games (radars, 3D graphics). Expertise was balanced within the participants.

[¶] ARTTracker1 is a product of A.R.T.: `http://www.ar-tracking.de/`

5.3.3 Controlled laboratory study 1

In the first study, we evaluate how different shapes impact on the performance of the pointing task, assuming that no panorama is available (as in Figure 5.20).

Independent variable. We had two independent variables:

- *Shape.* **Frontal**, **Bird's Eye** and **Top-Down**.
- *Angle.* For the annotations, we selected **12 distinct angular intervals** of size 30°, starting from [-15°, 15°].

Hypothesis. We expected a difference in time: Top-Down $<$ Frontal and Bird's Eye $<$ Frontal, because Top-Down and Bird's Eye, in contrast to Frontal, provide a direct representation of the mapping between panorama and environment.

Procedure. Participants performed 3 (Shape) × 12 (Angle) × 5 (repetitions) trials, for a total of 180 measurements per participant. We used balanced Latin squares to counterbalance the order of Shape and Angle. For each repetition, the crosshair was assigned a random yaw angle within the given angular interval. Since we did not use a panoramic image, participants had no reference point available for understanding how pitch angle mapped to the visualisation, and we could therefore not expect them to point accurately on the vertical direction. We therefore always assigned the crosshair a constant pitch angle of 0°.

First, we introduced participants to the three experimental conditions. We clearly indicated to all participants the front, left, right and back lines in the visualisations. Participants were allowed 6 practice trials for each Shape. We instructed all participants to be as fast and accurate as possible, but to give more importance to accuracy. Finally, we conducted a short interview, asking participants to describe the strategy they used to complete the task with each Shape.

Apparatus. We used the same setup as in the pilot study.

Participants. We recruited twelve university students and staff (6 males and 6 females), aged between 16 and 44 years (M = 29.5, SD = 7.4). All participants performed the experimental task using their dominant hand. All participants were not involved in our previous pilot study.

Results. For each of the 180 repetitions, we recorded the task completion time (in seconds) and unsigned error in yaw (in degrees) between the target and the selection. For each Shape × Angle condition and for each participant, we calculated the median time and error of the 5 repetitions.

Mean error values were 13.9° (SD = 4.0) for Frontal, 14.1° (SD = 1.4) for Bird's Eye, and 13.4° (SD = 2.6) for Top-Down. A Friedman test did not show any effect of shape on error ($\chi^2 = 3.5$, df = 2, p = .17). In the following, we analyse time measurements under the assumption of comparable accuracy.

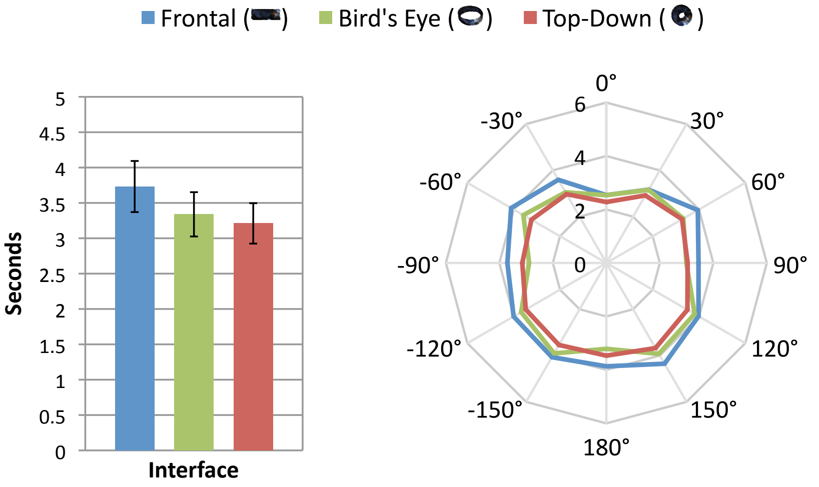


Figure 5.22: Results from user study 1. Left: the mean task completion time for each Shape. Right: the mean task completion time for each Shape × Angle (distance from center = seconds).

The mean task completion time (Figure 5.22, left) was 3.7 seconds (SD = 1.3) for Frontal, 3.3 seconds (SD = 1.1) for Bird's Eye, and 3.2 seconds (SD = 1.0) for Top-Down. Since the data violates normality and sphericity (ANOVA requirements) we conducted a non-parametric Friedman test, which revealed a significant effect of shape on time ($\chi^2 = 16.67$, df = 2, $p < .001$). Post-hoc Wilcoxon Signed Ranks tests with Bonferroni correction showed that all pair-wise differences were significant: Top-Down was significantly faster than Bird's Eye (p = .005) and Frontal (p = .003), and Bird's Eye was significantly faster than Frontal (p = .015). Bird's Eye was on average 3.5% slower than Top-Down; Frontal was on average 16.3% slower than Top-Down and 12.3% slower than Bird's Eye. In Figure 5.22 (right), we can see that Frontal was generally slower than Top-Down and Bird's Eye, besides for target locations around 0°. A Friedman test shows no effect of repetition on time, highlighting no significant learning effect.

In the interviews, participants reported that their principal strategy consisted of orientating in function of body-aligned directions (left, right, front, back) using the grid lines and the wind rose, and then refining the pointing direction. Participants reported imagining themselves “in the middle of the visualisation” for Top-Down and Bird’s Eye. For Frontal, more “thinking” with respect to body-alignment was required.

Discussion. Our results support our initial hypothesis, and highlight a further significant difference between the Top-Down and Bird’s Eye views. In general, we see that if no panorama is available users rely on body-aligned reference lines to orient themselves and refine the orientation of annotations between these lines. Shapes which correctly represent the spatial mapping of the panorama to the environment result in significantly shorter task completion times.

5.3.4 Controlled laboratory study 2

In the second study we evaluate how user strategies and performance change when a panorama is available.

Hypothesis. We expected a difference in time: Top-Down < Bird’s Eye and Frontal < Bird’s Eye, because we expected good readability to be advantageous for completing the task, and more advantageous than the strategy used in user study 1.

Procedure. We used the same setup and design as in user study 1. However, in this study we used a panorama of the room in which the study took place (a quiet laboratory room, shown in Figure 5.19). The panorama was up-to-date with the room’s appearance during the study. The panorama covers 360° in yaw, and [-45°, 45°] in pitch (in most real-world cases little information is present outside such range). We again gave the crosshair a constant pitch angle of 0°, to obtain results comparable with the ones from user study 1.

Participants. 12 university students and staff (6 male and 6 female), aged between 17 and 34 years (M = 24.4, SD = 6.5). None of the participants had taken part in the previous study.

Results. In this experiment, we calculated error as the great-circle distance (to include error in pitch) in degrees between the target and the selection. Mean error was 16.5° (SD = 8.4) for Frontal, 17.1° (SD = 8.0) for Bird’s Eye, and 15.7° (SD = 7.6) for Top-Down. A Friedman test showed no significant effect of shape on error (χ^2 = 4.2, df = 2, p = .12).

The mean task completion time (Figure 5.23, left) was 2.4 seconds (SD = 0.7) for

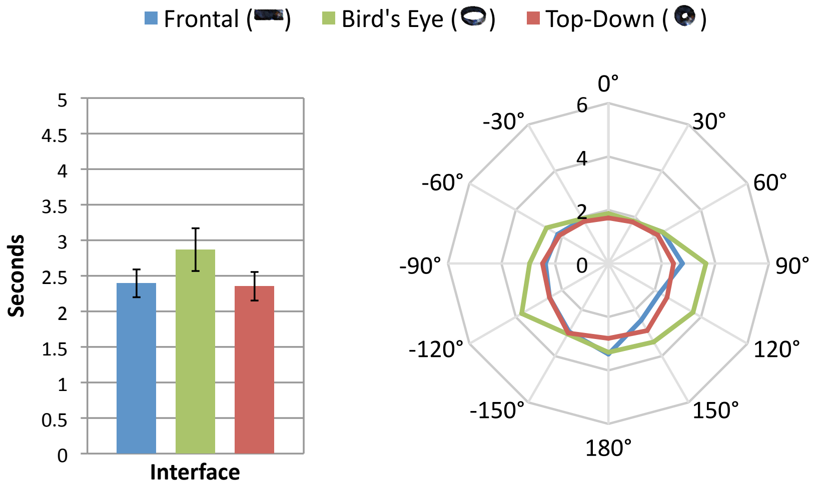


Figure 5.23: Results from user study 2. Left: the mean task completion time for each Shape. Right: the mean task completion time for each Shape × Angle (distance from center = seconds).

Frontal, 2.9 seconds (SD = 1.0) for Bird's Eye, and 2.4 seconds (SD = 0.7) for Top-Down. A Friedman test revealed a significant effect of shape on time ($\chi^2 = 20.67$, df = 2, $p < .001$). Post-hoc Wilcoxon Signed Ranks tests with Bonferroni correction showed that Bird's Eye was significantly slower than Top-Down ($p = .002$) and Frontal ($p = .002$). Bird's Eye was on average 21.9% slower than Top-Down, and 19.8% slower than Frontal. Friedman tests showed an effect of repetition on time, for all shapes ($p < .001$). However, Wilcoxon Signed Ranks tests only highlighted a significant learning effect between the first repetition and the others. This is not a problem for our analysis, since we perform it using the median values of the five repetitions.

In the interviews, participants reported that their strategy consisted mainly of looking for objects in the panorama and corresponding objects in the room (for example, a whiteboard and a fire extinguisher). For this strategy, most participants reported that Bird's Eye is inconvenient, as the sides are either not visible or warped and hard to see. One participant also reported issues with the backside of Bird's Eye, where the panorama appears mirrored. Figure 5.23 (right) illustrates this: we can see that Bird's Eye was slower than Top-Down and Frontal mostly around ±90°, in the warped or occluded regions.

Discussion. The results show that users adopt a strategy predominantly based on visual matching when a panorama is available, *independently of the shape.* As we hypothesised, Top-Down and Frontal were significantly faster than Bird's Eye.

5.3.5 Controlled laboratory study 3

In the third study, we aimed at corroborating the results from study 2 in the case of varying pitch angle.

Experimental design. Same as in study 2. In this study, the crosshair was assigned random pitch within [-45°, 45°] (the interval covered by the panoramic image).

Hypothesis. Same as in study 2: we expected a significant difference in task completion time: Top-Down $<$ Bird's Eye and Frontal $<$ Bird's Eye. This is because we expected visual matching to work also for non-zero pitch angles.

Participants. 12 university students and staff (6 males and 6 females), aged between 23 and 50 years (M = 30.1, SD = 7.8). None of the participants had taken part in the previous studies.

Results. Mean error was 15.0° (SD = 5.2) for Frontal, 18.5° (SD = 4.3) for Bird's Eye, and 16.6° (SD = 4.7) for Top-Down. A Friedman test showed a significant main effect of shape on error ($\chi^2 = 12.17$, df = 2, p = .002). A post-hoc Wilcoxon test with Bonferroni correction showed a significant difference only between Frontal and Bird's Eye (Z = -2.98, p $<$ .01).

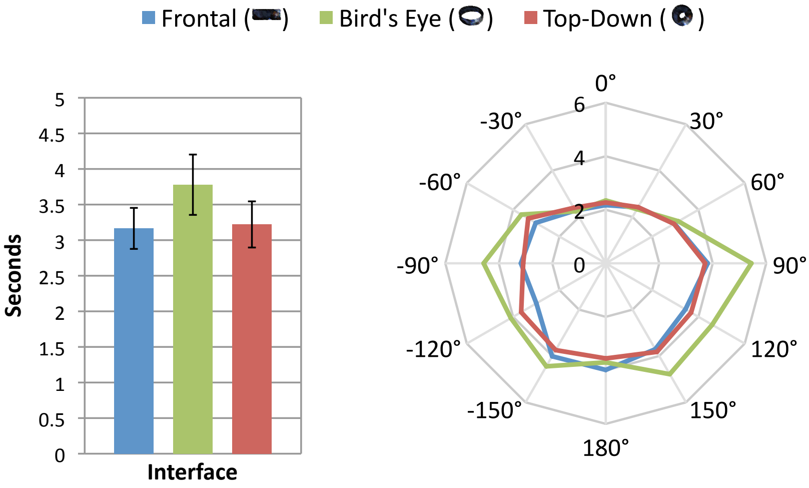


Figure 5.24: Results from user study 3. Left: the mean task completion time for each Shape. Right: the mean task completion time for each Shape × Angle (distance from center = seconds).

The mean task completion time (Figure 5.24, left) was 3.2 seconds (SD = 1.0) for Frontal, 3.8 seconds (SD = 1.5) for Bird's Eye, and 3.2 seconds (SD = 1.1) for Top-Down.

A Friedman test revealed a significant effect of shape on time ($\chi^2 = 15.17$, df = 2, p = .001). Post-hoc tests showed that Bird's Eye was significantly slower than Top-Down (p = .002) and Frontal (p = .003). Bird's Eye was on average 16.1% slower than Top-Down, and 17.1% slower than Frontal. A Friedman test showed no effect of repetition on time.

In the interviews, the strategies and issues with Bird's Eye reported by participants were in line with the strategies reported by the participants of user study 2. Figure 5.24 (right) illustrates that Bird's Eye was again slower than Top-Down and Frontal mostly in the warped areas. Participants reported that Frontal was the easiest condition for finding the pitch of points, whereas finding the yaw was considered harder than with the other conditions. This was particularly true for the annotations to the sides and the back of participants, because they required more thinking to resolve the mapping from the rectangular shape to the environment. However, this was a subjective consideration and a slowdown was not observed in our quantitative measurements. With Top-Down, participants had issues with the outermost third of the visualisation (high pitch in the panorama), due to the strong distortion effect.

Discussion. The results of this study replicate and confirm the results of the previous study, also for the case of varying pitch angle.

5.3.6 Discussion

Our results show that, in the presence of a panorama, users mainly perform visual matching of objects in the panorama with correspondences in the real environment. Consequently, good readability of the panorama has primary importance in the design of a panoramic overview. As a secondary strategy – and the main one when no panorama is available – users look for body-aligned directions within the visualisation.

Applying these results to location-based services and AR, we see two advantages in the adoption of Top-Down. First, it provides good readability of the visual elements in the panorama, and our results show that this is what users need the most. Second, since we typically have incomplete panoramas (see Figure 5.11) or pre-recorded panoramas that do not match the current appearance of the environment, Top-Down is designed to clearly represent the mapping between panorama and environment and therefore provides a valuable fallback when visual matching between panorama and environment is not possible.

5.4 Reflection on our research questions

If we consider all results of this chapter from the perspective of our main research questions (Section 1.3), we infer that an AR technique like Tunnel fails in communicating the spatial position of specific pieces of information correctly, as it gives a weak spatial awareness while imposing relatively high workload on users. In contrast, combining AR with Mosaic and Transitional is a more effective solution for pointing users towards a specific piece of information, and for supporting the understanding of the reciprocal position of the user and the information. Our results therefore suggest enhancing AR with compass-like overlays or transitional interfaces to improve users' awareness of the position of surrounding information.

The good results with Mosaic suggest that AR succeeds in supporting users in pointing tasks, provided that extra overlay information is given to guide the pointing. These results are confirmed by our second experiment, which shows that the information provided by Compass is sufficient if the system can highlight the target object: in real-world applications this is a case analogous to the one presented in Section 5.1, and is applicable if the system knows what the user is looking for (for example, if the system must highlight the best result of a search query).

In contrast to this, AR fails in providing sufficient overview for more complex search tasks. In these tasks, however, zooming interfaces are an effective complement to AR. The zooming interfaces support users better in exploring the information, for all tasks where a more integrated view is needed. This would rather be the case when multiple options are available, and the selection criteria of a user are not clear or not easily described via a software interface.

In summary, the results from our experiments suggest a number of design recommendations. First, integrating compass-like overlays into AR navigation systems is an effective way to aid people in pointing towards specific augmentations. Clearly, this depends on the ability of the system to infer what the user is looking for and to highlight the appropriate augmentation that fits the user's needs. Second, transitioning to other interfaces is an effective way to provide overview of all information, when people need it. Maps and panoramic overviews are both good at providing this overview, and they are probably complementary interfaces: when people need overview of distant augmentations, maps are probably a better choice because they are designed to show information at several blocks of distance from the user; when people need overview of augmentations within line of sight, panoramas might be more helpful than maps, because they provide a first-person view of the surroundings. Finally, with respect to panoramic overviews, a circular shape that communicates both directional information and the visual landmarks in the panorama is probably a better design choice than a rectangular panorama, because it communicates more clearly physical directions in the environment.

Chapter 6

Exocentric exploration

In this chapter, we focus on exploring information with *exocentric views.* In contrast to the previous chapter, in this case we augment the information on a paper map, rather than visualising it directly at its corresponding physical location in the environment. Users can browse the information in a natural way by panning the phone over the map.

Augmenting a paper map provides users with a top-down exocentric point of view on the information, which gives a broader overview on the information than egocentric exploration. However, in this case information is not directly visualised at corresponding locations in the environment: in order to understand the physical location of information users must mentally resolve the transformation from the map to the environment. An augmented map therefore provides a support complementary to the solutions discussed in the previous chapter, as it allows for a broader, exocentric point of view on the information, but it does not anchor the information to the environment as directly as in the case of egocentric exploration.

This chapter investigates *how users exploit augmented maps during navigation.* The work presented in this chapter is partially based on previous results by Morrison et al. [116, 118], to which the author did not contribute. We briefly present these previous results in the next section, to give the reader a clearer understanding of the context of our work.

6.1 Previous results

MapLens is an AR application developed for Nokia camera phones (Symbian OS S60) with GPS. When users point their device at the paper map, geo-referenced photographs are retrieved from an online database and augmented on the map.

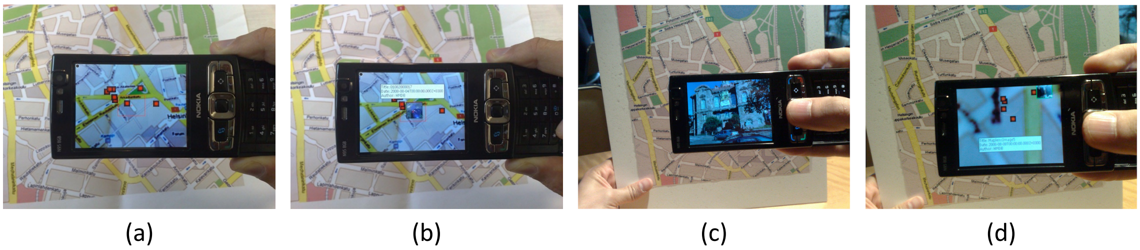


Figure 6.1: The first prototype of MapLens. Typical interaction with the information: (a) browsing images on the map, (b) hovering the phone over an icon shows a thumbnail of the corresponding image, (c) clicking on a thumbnail shows a full-screen version of the image, (d) when icons are cluttered, users can freeze the video view and magnify it.

Interface design. The first prototype of MapLens (Figure 6.1) overlays the map with red icons that identify geo-referenced images. Hovering over an icon shows a thumbnail of the related image. When a thumbnail is visible users can click on it to see a full-screen version of the image. A freeze function helps the selection when multiple icons are close together: when a user clicks over more than one icon, the view is frozen and magnified so that users can more easily select a specific icon.

MapLens also functions as a photo camera. The user can press the * key to enter the camera mode, **0** to capture a photo, and * again to return to MapLens. Photos are automatically geo-tagged and uploaded to the online database. All other MapLens users receive the new photo within a few minutes. By pressing **1**, one can see photos taken by other users. Pressing **1** again turns that layer off.

The first prototype of MapLens operates at 5–12 frames per second, allowing for interactive use. Operation is possible within a distance of 15–40 cm between the printed map and the camera. The camera can be tilted relative to the map up to a range of ±30 degrees from the perpendicular view. In-plane rotation (around the viewing axis) is handled over 360 degrees. MapLens operates on A3 printouts of Google Maps (street layer).

A purely digital map phone application called *DigiMap* was also implemented as a comparative interface. DigiMap uses a digital version of the map used for MapLens. Like in MapLens, red icons indicate geo-referenced photographs. Users scroll the map with the phone joystick and zoom in and out with two buttons. DigiMap does not access the phone's camera, so users must switch to the native camera application to take photographs.

User study. Morrison et al. [118] conducted an evaluation to compare MapLens and DigiMap usage. Out of 37 participants, 24 shared MapLens in 9 teams, 2 used MapLens solo and 11 used DigiMap (5 teams). The experiment was run as a location-based treasurehunt game in the city of Helsinki, designed to promote awareness of local environmental issues. The game began at the Natural History Museum where players completed indoor tasks, which served as a "warm-up" for teams to get to know each other. The game then continued outdoors. Overall, players solved a variety of types of tasks (14 in all, see Figure 6.2), some of which were sequential problem chains.

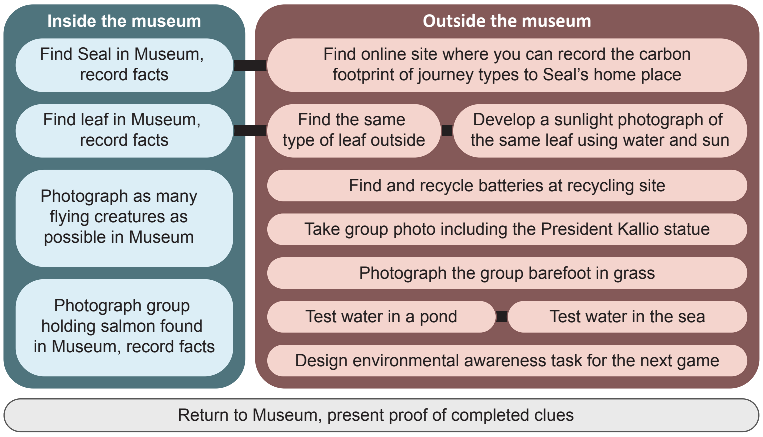

Inside the museum
Find Seal in Museum, record facts
Find leaf in Museum, record facts
Photograph as many flying creatures as possible in Museum
Photograph group holding salmon found in Museum, record facts
Outside the museum
Find online site where you can record the carbon footprint of journey types to Seal's home place
Find the same type of leaf outside
Develop a sunlight photograph of the same leaf using water and sun
Find and recycle batteries at recycling site
Take group photo including the President Kallio statue
Photograph the group barefoot in grass
Test water in a pond
Test water in the sea
Design environmental awareness task for the next game
Return to Museum, present proof of completed clues

Figure 6.2: Tasks for the game used for the experiment. Interconnections indicate sequential tasks.

How tasks were completed and in what order was up to the players. Some tasks could be completed in several places, whereas a series of tasks required visiting specific places in a certain order. Using MapLens, players could access clue images, as well as the photographs taken by the other participants. Players therefore required the assistance of the technology to follow the clues necessary for completing the game tasks. Each team worked with a kit that contained seven objects (see Figure 6.3), including one mobile phone and one paper map.

The experiment was exploratory, looking at how MapLens is used by the players, compared to DigiMap. Mainly qualitative methods were employed to analyse player behaviour. Each team was accompanied throughout the game by one researcher taking notes, photographs and videos. Participants also described their experience in semi-structured interviews, highlighting aspects that had caught their attention. Throughout the trial, participants took photos as evidence of completing tasks. These images assisted researchers to build an overview of activities undertaken during the trial.

Results. MapLens players showed stronger *collaborative behaviours* than the DigiMap players. Stopping and gathering around the paper map created an opportunity to focus on a problem as a team: the physical map acts as a place where joint understanding

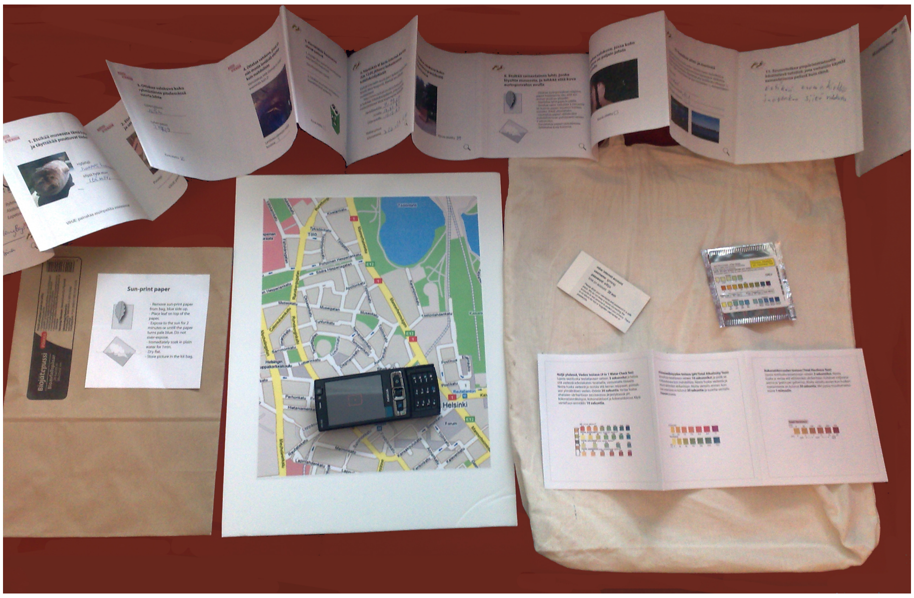


Figure 6.3: The game kit used for the experiment. Each team was given a kit that contained 7 items: sunlight photographs, map, phone, water testing kits, voucher for Internet use, clue booklet and pen.

can be reached. Yet, the AR technology of MapLens also restricted players' movements, as they had to stabilise the physical map and the device to be able to operate the system (see Figure 6.4, left). MapLens players consequently favoured places where they could place the map on a table, on a bench, or on the ground, or they held the map for the other team members. Establishing common ground was easier for MapLens players than for DigiMap players, as the position of the device on the paper map and the contents on its display helped players understanding the points under discussion without explicitly needing to ask. The map-device combination triggered collaboration in a physical way using fingers, pens and other objects. However, some MapLens players found it challenging to identify the location on the map through the screen of the device, especially while the device was in use by another player. MapLens players also handed over the phone to other team members more often than the DigiMap players.

In contrast, DigiMap players only needed to stop at places that the tasks themselves dictated. DigiMap players typically kept the device lower and closer to the body: this posture made the phone more private as others could not see the contents on the display (see Figure 6.4, right). DigiMap teams were also not able to share the map easily and often referred more directly by pointing at their surroundings. Typically, in DigiMap teams one person took the role of "navigator" and used the device solo.

Seven of the eleven MapLens teams tried using the system while walking, but all faced

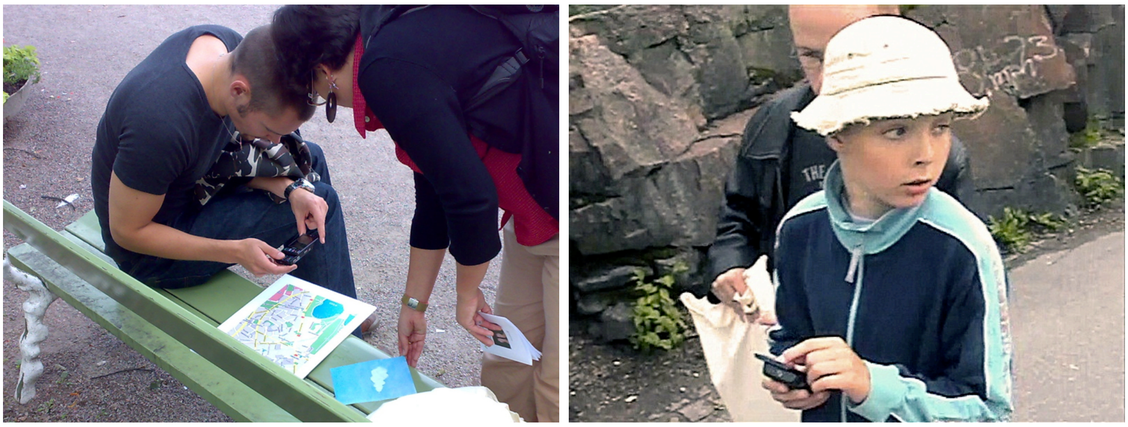


Figure 6.4: Pictures from the first MapLens experiment. (Left) Stopping, place-making and sharing with MapLens often required laying the map on a stable surface. (Right) DigiMap was not easily shared.

difficulties. First, even a light trembling of the device hinders MapLens usage because the AR tracking fails. Second, awareness of the environment is challenged when using MapLens (e.g., one player walked into a lamp-post). Overall, MapLens did not support usage while walking. Difficulties with use while walking were not as common for DigiMap players. Three teams used the system while walking, and one team even ran while watching the map.

On the other hand, the results show that the AR interaction in MapLens and the availability of a paper map foster collaborative behaviours that are not present in a digital-map application, which is a more personal interface and harder to share. On the other hand, the necessity to synchronise device and map in order to operate MapLens, and the limitations of the tracking technology, inhibit spontaneous usage in some situations, for example while walking or when there is no stable surface nearby.

6.2 Refined interface design

We redesigned MapLens (see Figure 6.5) taking into account the main findings from the previous experiment. First, we improved the tracking technology to support more spontaneous usage of MapLens. The second prototype runs at 16–20 frames per second, allowing for a much smoother interactive experience. The improved tracking technology (see Chapter 3) is robust to changes in illumination (sunlight), blur in the camera image and allows for camera viewing tilts of up to almost 90 degrees. The new tracking technology also supports camera distances from the map between 10 cm and 2 m.

Second, we tried to further support usage while on the move in our design. We added

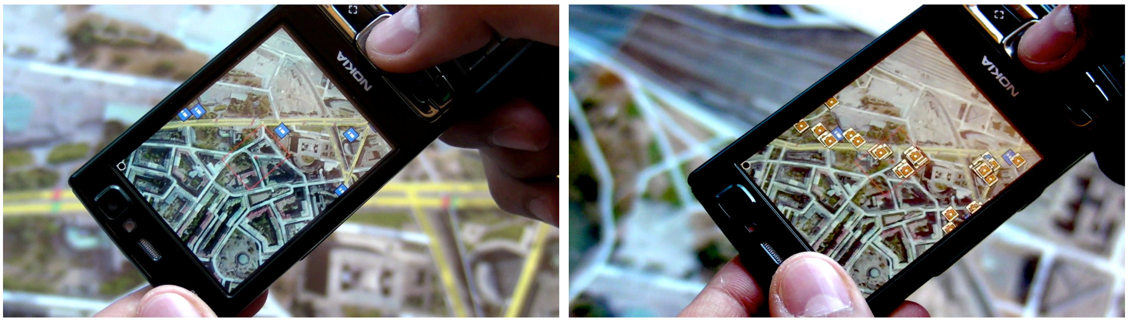


Figure 6.5: The second MapLens prototype. Left: browsing the map, when only clue icons are enabled. Right: browsing the map after user-generated photographs have been enabled.

a thumbnail bar, shown in Figure 6.6. When users select a specific area of the map, the thumbnail bar shows an array of all preview images available at that location. Green and dashed visual links help maintaining the connection between the thumbnails and their location on the paper map, without interrupting the AR experience. The thumbnail bar is independent from the AR interface: once the thumbnail bar is active, users can remove the paper map and still browse the chosen images. As soon as the paper map is in view again, the visual links reappear. This design therefore allows two usage modes: a first one with AR on the paper map, a second one without the paper map and without AR. The latter mode has less functionalities (users can only browse the images), but it also requires less motor coordination to be operated, as there is no need to synchronise phone and map.

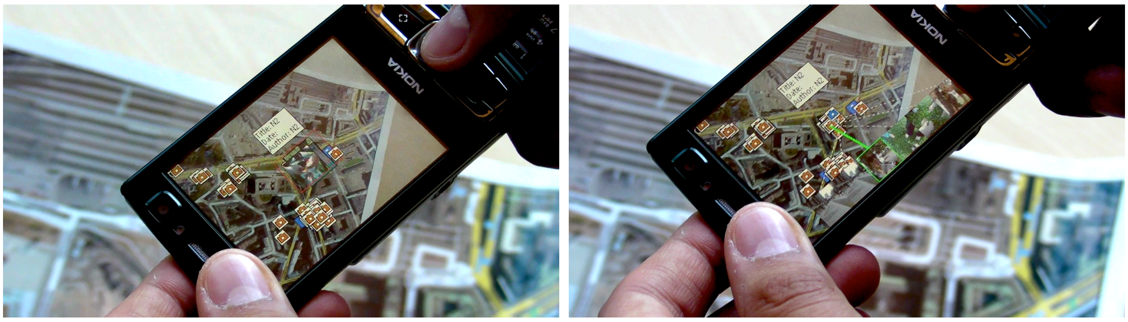


Figure 6.6: Typical interaction with the information in the second MapLens prototype. The user is browsing image thumbnails in a specific area (left). After selecting the area, a thumbnail bar (right) appears. All image thumbnails are now accessible as a single array of images visually linked to their corresponding location on the map.

Finally, we also addressed a few smaller usability issues. We replaced the keypad combination needed to capture a photograph with a single long press of the camera button built in the phone: this solution has a more straightforward affordance than the previous keypad-based approach. We added a "you are here" icon to show the position of the user on the map, an important piece of information that was missing in the previous design. As shown in Figure 6.5, we visualise pre-determined game clues and user-generated photographs with different icons, to prevent misunderstandings. The new prototype operates on A1 printouts of Google Maps satellite images with street overlays.

6.3 Exploratory study

We conducted one user study with the refined prototype of MapLens. We designed the study with exploratory goals: first, we wanted to confirm that the refined prototype also fosters collaborative behaviour. In the previous evaluation there were only one map and one shared device per team. This might have forced players to collaborate in order to use MapLens. In this experiment, we eradicated this factor by giving each player a separate map. Further, in some teams we gave each player one phone, and investigated multi-device use to see if it changes the way people collaborate. This is a more realistic scenario, as in everyday life users will typically have their own device, which they will want to use for browsing the augmented map. Finally, we wanted to verify if the improved tracking and the novel thumbnail bar allow for a more spontaneous usage while on the move.

Independent variable. We compared teams sharing one single phone (*single-device*), to teams in which each player was given one phone (*multi-device*), and to *solo* users (as a control condition, to verify that solo usage is possible).

Procedure. To ensure that results were comparable to the previous ones, we adopted the same experimental design and outdoor game as in the previous evaluation.

Participants. 37 people (19 females and 18 males), aged between 14 and 44, participated to the user study. Participants were largely professionals with university qualifications, early-adopters, and/or researchers. 21 participants were divided into 7 multi-device teams and 12 participants were divided into 4 single-device teams. All teams were composed of 3 players. Finally, 4 participants played the game solo.

Results. All solo users completed the game, therefore solo usage of MapLens is possible. In the team conditions, we could observe collaborative behaviours analogous to the previous experiment (see Figure 6.7): players gathered around the map and used the opportunity to discuss problems and to create common understanding within the team.

Shared map and device. We observed that the availability of multiple devices has an impact on the way players collaborate. Single-device players seem to communicate more

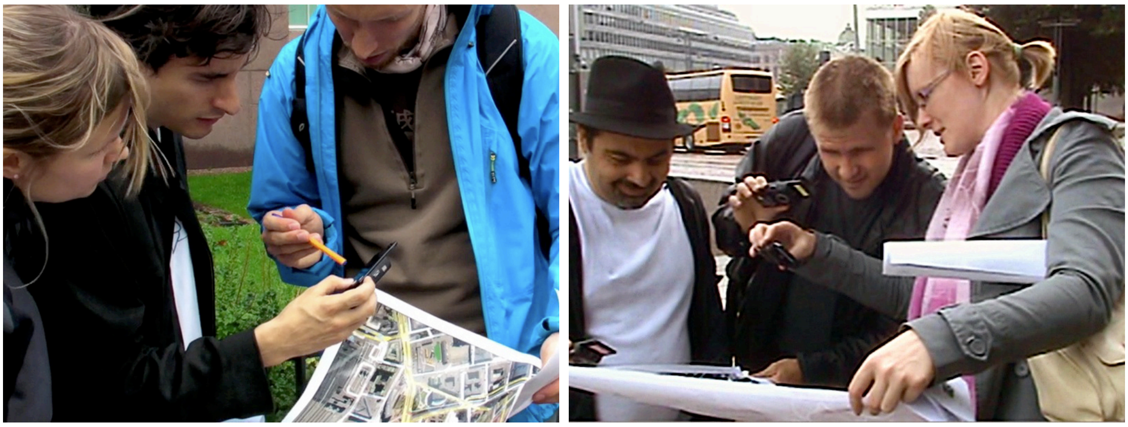


Figure 6.7: Establishing common ground in single-device and multi-device teams. (Left) Single-device players communicate and establish common ground more around the system. (Right) Multi-device players establish common ground more through the system. Here, the players are using two devices, and keeping them on different heights to avoid collisions.

around the system by sharing information on the map, screen and environment (Figure 6.7, left). In contrast, multi-device players seem to communicate less with each other, and to share information more looking through their own devices (Figure 6.7, right).

We looked for support to these observations from our video recordings. We cut video footage into manageable chunks, focusing on players' activity around MapLens. The footage from two teams (1 single-device, 1 solo) was excluded from the analysis due to technical failure. We employed a team of four researchers to code players' actions in our video recordings. Through inter-researcher agreement, we compiled an initial 52-item list of tasks performed by players. We iterated the list over several sessions until we isolated and identified four principal game tasks:

- *Device usage.* The player is using the device.
- *Map usage.* The player is carrying, orienting or holding out the paper map.
- *Navigation.* The player is discussing navigational decisions.
- *Scouting.* The player is exploring the environment, using the clue booklet, taking photos, discussing with other players.

We coded all occurrences of the four tasks in our recordings. It must be stressed that, due to the small sample size (7 multi-device and 4 single-device teams) we cannot conduct a statistical analysis of this data, but we rather analyse it qualitatively as a support for the

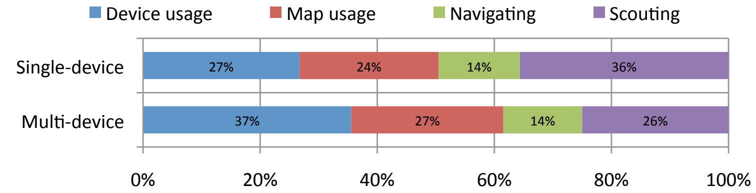


Figure 6.8: Division of the four principal tasks in single-device and multi-device teams.

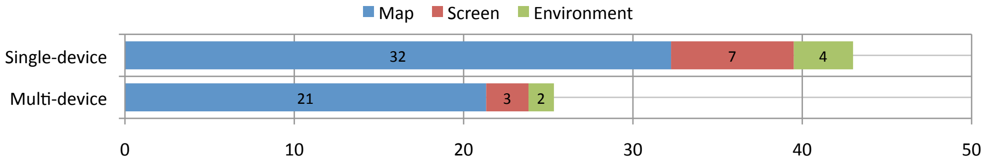


Figure 6.9: Communication in single-device and multi-device teams. We present communication as average occurrences of pointing at the map, pointing at the device's screen and pointing at the environment.

observations we made during the experiment. Figure 6.8 shows that, on average, multi-device teams used the device more than single-device teams, who in contrast engaged more in scouting activities. This suggests that multi-device teams worked more through the device, while single-device teams performed more activities outside the device. We further coded all occurrences of communication between players during device usage (see Figure 6.9). We observed that pointing to the screen of another team member (see Figure 6.7, left) occurred more often in single-device teams. A similar effect is visible, to a smaller extent, in the amount of pointing to the map and to the environment. During the experiment and in the video recordings, we observed that single-device players shared the device screen largely throughout all sessions, by tilting their screen for others to see, pushing the device closer to others, handing the device over and standing closer together. In multi-device teams, we observed that the intentional sharing of screens happened less, typically only a couple of times during the game, and only for a few seconds. Our results show that there was a trend for multi-device players to focus more on the device (see Figure 6.7, right) and less on communication, as they could all synchronously experience the same AR view of the information. In contrast, single-device players seemed to communicate more with each other than multi-device players: this extra amount of communication was probably necessary for them to reach the same level of shared understanding of the game status as multi-device players.

When starting the game, multi-device teams typically used two or three devices simultaneously. However, as the game progressed one main device (mean total use 51%)

and secondary/tertiary devices (mean use 33% and 16%) emerged. Similarly, two or three maps were used in the early training stages, but teams quickly switched to one main shared map. Two phones seemed to be the maximum amount of devices that could simultaneously fit over a map of this size: devices were used in a panning motion over the map and needed space around them in order to move freely. When devices collided, one user moved the device to a different height above the map (as in Figure 6.7, right), moved alongside the other device on the map, or withdrew the device and looked through the other device. Some players used multiple devices to simultaneously explore different areas of the map.

Flexibility and usage on the move. Teams across all conditions used the system not only after setting down the map on a supporting surface but also while standing and holding the map (like, for example, both teams in Figure 6.7). Three of the 15 teams (2 single-device, 1 solo) only used MapLens while standing and never put the map down onto some surface. With the improved technology of MapLens, we observed that this agile *parking* behaviour emerged (briefly stopping to check a detail before moving on). This is a new behaviour not observed in the previous experiment, in which *stopping* (standing for longer periods of time or setting down the map) was the only mode of usage.

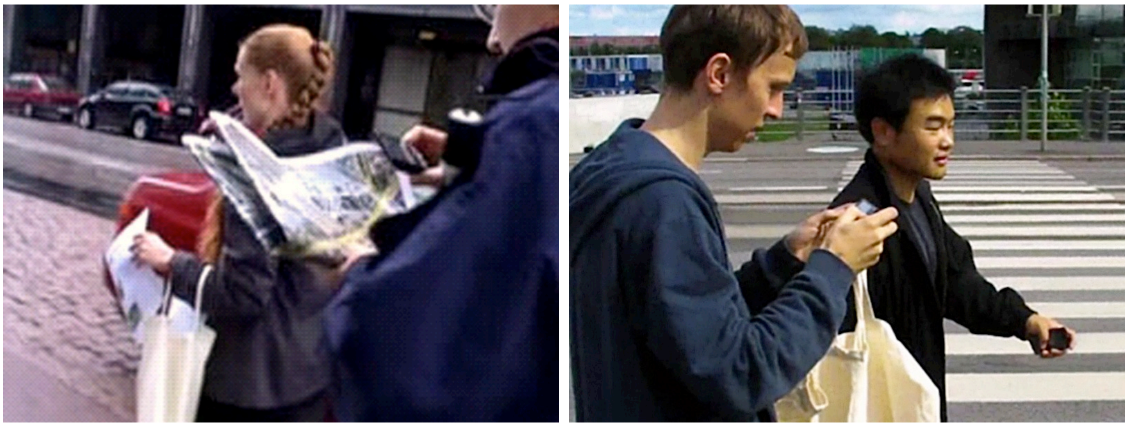


Figure 6.10: Using MapLens while walking. (Left) Walking and using MapLens was possible with a folded map. (Right) Walking and browsing photos in MapLens did not require usage of the paper map.

We also observed that usage while walking was possible with the improved technology (Figure 6.10, left) as there was no need to shade the display for tracking to work, or to keep a steady hand while standing still. However, the need to pay attention to the surroundings while reading the display prevented systematic usage while walking. Rather, we observed that players mostly exhibited the aforementioned parking behaviour, for using MapLens while on the move. Finally, we observed that the thumbnail bar allows exploration of information while walking (see Figure 6.10, right) and reduces the need for recurrent stopping. In general, we noticed that this feature supports more agile forms of using the system.

In post-experiment interviews, people reported being very engaged and involved with the game, although several users reported that the most engaging part of the experience was the technology. Almost all users reported having used pointing to the map as a means of communication between team members, and half of them reported that pointing was very helpful to refer to items seen through MapLens. Two participants mentioned augmenting public maps as an interesting further development of the project – one of these two participants stated that he would not use MapLens if he had to carry the map with himself all the time, but he would certainly use it if it worked with public maps.

6.4 Reflection on our research questions

If we consider the results of this chapter from the perspective of our main research questions (Section 1.3), we can see that augmenting a physical map allows easy communication, collaboration and information sharing between users, in contrast to a digital map which is a more private and personal interface. Users are quickly able to grasp the interaction metaphor of browsing the augmented map by moving the camera phone over it. Given robust tracking technology, users are able to browse information on augmented maps in spontaneous and flexible ways, from setting down the map on the ground and discussing with other users, to briefly stopping on the sidewalk for a quick glance at the information. While users often share their device with each other, the availability of one device for each user allows sharing the AR experience more effectively, reducing the communication effort needed by users to create a common understanding on the information.

Operating an augmented map is hard while walking. Even if robust tracking technology can support usage while walking, the need to synchronously move map and device requires bimanual coordination, which users seem to find too demanding while walking. In general, users typically prefer shortly stopping to browse the information.

While AR is needed to browse the information on the augmented map, it is often not necessary when users want to explore the details of a specific set of augmentations. For usage while walking, or in other contexts where bimanual operation of the interface is too demanding, AR is better combined with a less demanding non-AR interface: the AR interface can then be used to select certain pieces of information from the map, which users can then explore in more detail in a non-AR mode.

PART IV

Wayfinding

Chapter 7

Outdoor navigation

In this part of the thesis, we focus on *wayfinding*: finding the way to reach a specific destination. During wayfinding, AR can be used to augment the environment with cues that support correct navigational decisions.

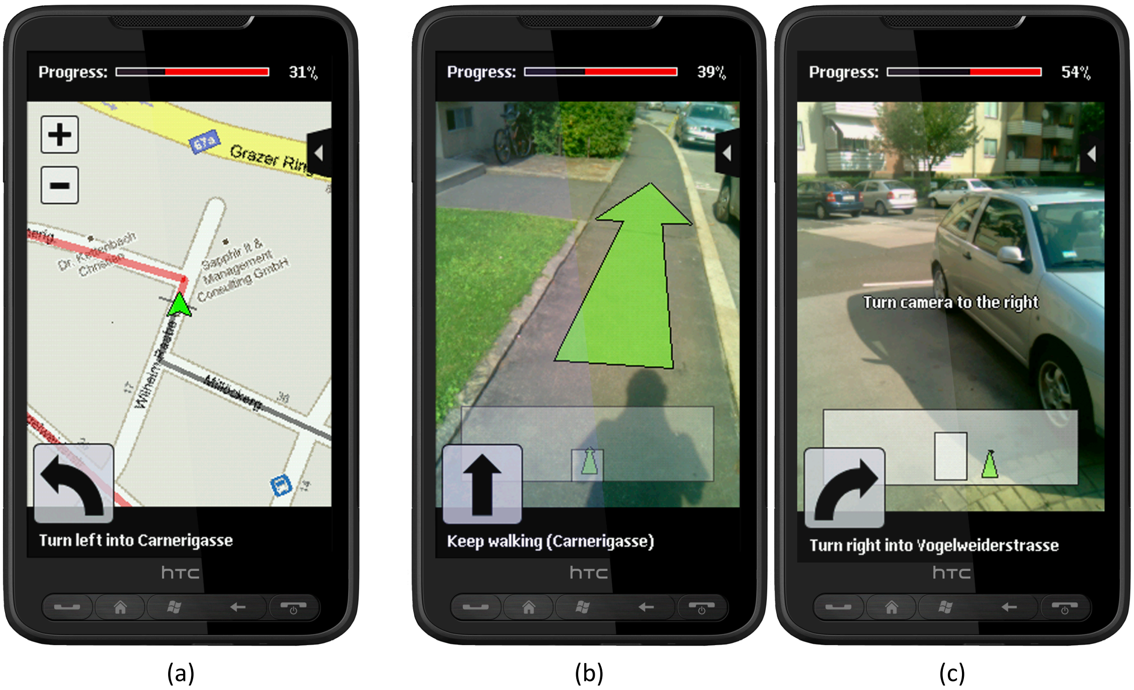
In this chapter, we target outdoor navigation. During outdoor navigation, people need to traverse a sequence of decisions points, where they have to make wayfinding decisions. In between such decision points, there are usually long moments of purely mechanical travelling, on path segments where people need little or no external support. It is therefore questionable that people will use AR continuously during navigation, while a more realistic assumption is that they will need AR to support their wayfinding only under specific circumstances.

This chapter investigates *scenarios in which users can benefit of AR during navigation.* To gain first insight on this, we present a preliminary evaluation of an outdoor navigation system enhanced with AR.

7.1 Interface design

We designed a multimodal navigation system (Figure 7.1) with a forward-up map highlighting the user's position and the path to be followed. Navigational hints are explicitly provided as glyphs. To support eye-free usage, audio instructions are provided and every new instruction is notified by the phone vibrating. From here on, we will refer to the combination of all these interface elements as the *map interface.*

In addition to the map interface, we provide users with an on-demand *AR interface.* We superimpose virtual arrows onto a live video stream of the physical world, indicating the direction the user should follow, as in other AR navigation systems. Similarly to the



Turn left into Carnerigasse

Keep walking (Carnerigasse)

Turn camera to the right

Turn right into Vogelweiderstrasse

(a) (b) (c)

Figure 7.1: Screenshots of our prototype. (a) We employ maps, glyphs and text instructions to provide information about the path and the next turn. (b) We augment the physical world with arrows to provide egocentric navigational cues. (c) When the augmentation is not visible, visual cues guide the user in pointing the camera to the right direction.

work presented in Section 5.2.3, tilting motions trigger transitions between map and AR: tilting the phone down shows the map, tilting it up transitions to an AR view. When an arrow is outside the view of the phone's camera, we overlay information to guide the user in turning the phone towards it, using the Compass interface presented in Section 5.2.3.

7.2 Exploratory study

We conducted an exploratory study of our system in a real-world navigational task. We looked closely at how the participants used the system, observing where, how and how often they exploited AR during navigation.

Task. Participants were asked to follow a predefined path of 1.67 kilometres in a residential area in the city of Graz, Austria (Figure 7.2, left) with the aid of our navigation system. We selected an area with relatively low buildings to provide good GPS position measurements: as shown in Figure 7.2 (right), throughout the study the GPS quality was excellent most of the time (Dilution of Precision (DoP) < 2, green), and only rarely good

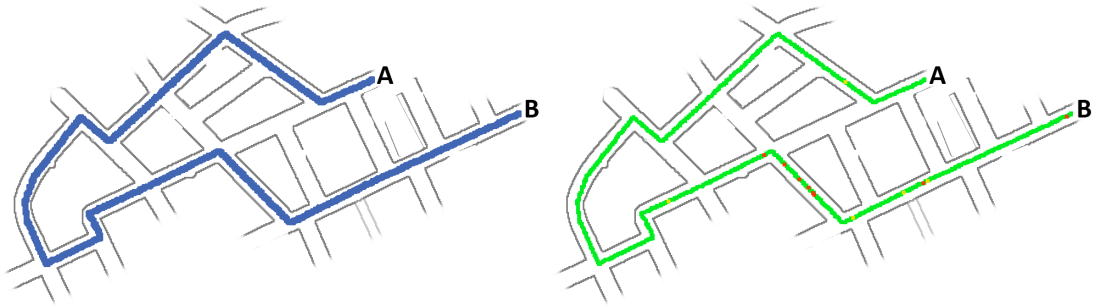


Figure 7.2: The path used for the user study. Left: participants were asked to follow a predefined path of 1.67 km from A to B. Right: average quality of the GPS signal on the path (green = excellent, yellow = good, red = moderate).

($2 < \text{DoP} < 5$, yellow) or moderate ($5 < \text{DoP}$, red). DoP is a measure of the geometric accuracy of the triangulation of GPS measurements from different satellites, and depends on the relative positions of the satellites and the GPS receiver. Values of DoP smaller than 2 are considered a level of accuracy sufficient for navigation systems [89].

Hypotheses. We aimed to observe *where*, *how* and *how often* users exploit AR during outdoor navigation. We hypothesised that the usage of AR, despite being available everywhere on the path, would be mainly clustered around decision points (road intersections) on the path. We also hypothesised that AR would be used mostly while standing still. The two hypotheses are in line with previous findings on the usage patterns that occur with augmented photographs [24, 58, 204].

Procedure. After briefing participants on the study modalities, we walked them to the starting point of the path while they practiced with the system. We reminded participants to follow the path (without taking shortcuts) and to be free in the usage of the device (without feeling forced to use the system continuously). We set up the study to gather rich information on how the system was used: we employed continuous software logging of the system status, and we followed and video recorded participants through the whole task. After the task was completed, we collected demographic and experience information through a questionnaire, and subjective feedback through a semi-structured interview.

Apparatus. The system runs at interactive frame rates on an HTC HD2, a smartphone with GPS, an accelerometer and a magnetometer. We track the position of the phone using GPS and its orientation using accelerometer and compass. We use a linear Kalman filter to remove high-frequency noise from the sensor measurements. It must be noted that at the time of the experiment sensor-based tracking reflected the state-of-the-art for commercial handheld navigation systems using AR. We also combine sensors and vision-based panorama tracking (Section 4.2) to obtain a more stable orientation tracking when the system is used while standing still. To orient the forward-up map in real time, we use magnetometer measurements. We designed the system for a phone in the upright position, to guarantee optimal grip for one-hand usage or with the arm in the rest position.

Participants. Nine people (p1–p9) participated in the experiment, aged from 25 to 33 (M = 28.1, SD = 2.6). Three participants were familiar with the area in which the study was conducted (p7, p8, p9). Five participants own and frequently use some navigation system (p1, p2, p4, p5, p8). Three participants had previously used an AR application (*AR experts*: p1, p2, p4); it must be stressed that these participants had no background in AR research or development: they were familiar with the AR interaction metaphor but were not knowledgeable of the functioning details of an AR system (for example, how tracking is implemented). The other participants had never used an AR application before (*AR novices*: p3, p5, p6, p7, p8, p9). None of the participants were familiar with AR navigation systems.

Results. All participants completed the task successfully. Since none of the participants exited the predefined path, no intervention from the evaluators was needed. We synchronised all video recordings with the corresponding software logs, and cut them into *usage sessions* – we defined a usage session as a time interval in which users were constantly looking at the device's screen. Based on the identified usage sessions, all corresponding segments from the software logs were then extracted for analysis.

Usage time analysis. Participants used the system on average 23.1% (SD = 11.9) of the overall task time. On average, 28.7% (SD = 22.5) of the system usage was on AR and the remaining 71.3% (SD = 22.5) on the map, the text instructions and the glyphs. The usage time of a session in which AR was accessed was 4.8 seconds (SD = 2.3), whereas sessions in which only the map was used lasted on average 1.8 seconds (SD = 0.6).

Figure 7.3 shows the percentage of time each participant spent using the system, distinguishing between usage of AR and usage of the map interface. AR experts used the system on average 28.8% (SD = 5.4) of the overall task time, and exploited AR for 16.2% (SD = 2.4) of the overall task time (more than half of the usage time, 57.1%, SD = 8.5). AR novices used the system on average 17.4% (SD = 13.4) of the overall task time, but they used AR only 1.9% (SD = 0.8) of the overall task time (14.5% of system usage, SD = 7.5). In contrast, map usage was similar between the two groups: AR experts used

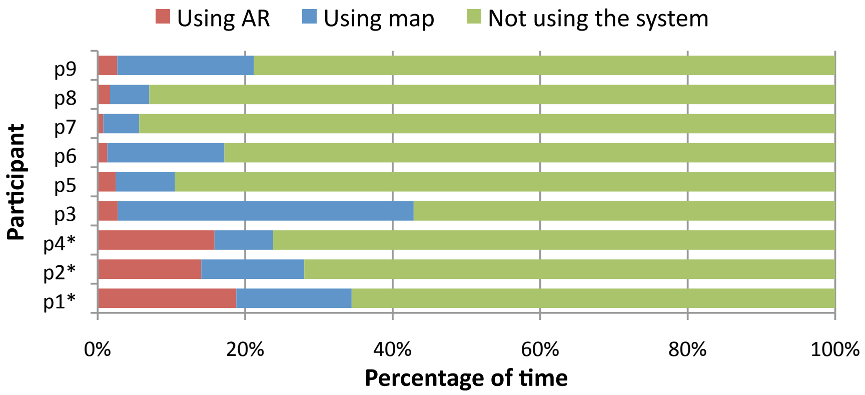
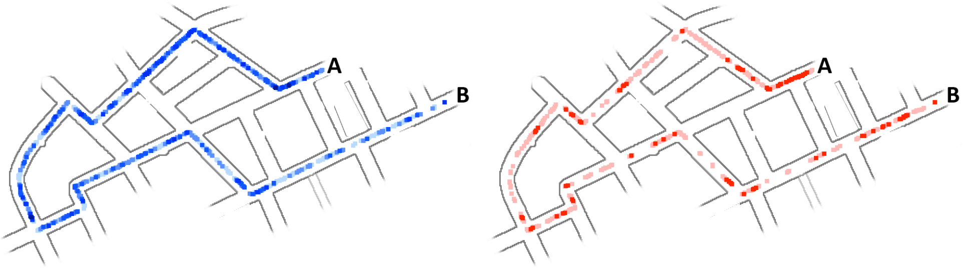


Figure 7.3: The percentage of time each participant spent using the system. We distinguish between usage of AR and usage of the map interface (* indicates participants with previous AR experience).

the map 12.5% of the overall task time (SD = 4.0), while AR novices used the map 15.5% of the overall task time (SD = 13.3) (if we exclude p3, whose map usage was larger than 3 times the inter-quartile range, the map was used 10.5% of the overall task time, SD = 6.2).

These observations suggest that there were differences in the amount of use of AR between AR experts and AR novices. In the interviews, AR experts said that the AR view was useful when it was not clear which street turn to take (p1, p2), or when the signs with the street names were not easily spotted (p4). AR novices commented that they did not use AR because the map information was sufficient (p6, p9), more familiar (p5) and gave a better overview of the path (p3), or because the visualisation of the AR arrow was not stable enough (p8).

Spatial analysis. Figure 7.4 shows where the participants used the AR and map interfaces on the path. While the map interface was used almost uniformly throughout the path, the AR interface was used less on straight path segments and more at road intersections. In Figure 7.5 we look in greater detail at where the participants used the AR and map interfaces on a smaller scale. For this, we define path segment as one section of the path comprised between two consecutive road intersections. It should be noted that the plots show aggregate data from only a small number of participants, and we therefore limit ourselves to a qualitative analysis. In the plots, we see that system usage generally increased when approaching a next intersection. For AR experts, we see that AR usage increases just before an intersection (decision on the turn to take) and shortly after it (confirmation of being on the correct street). This observation is in line with the feedback from the interviews on the usefulness of AR at path turn, as well as with our observations during the experiments and in the video recordings (see Figure 7.6). Map usage by AR



Map Augmented Reality

Figure 7.4: The locations in which the prototype was used. We distinguish between map (left, blue) and AR (right, red) interface use. The plots show the number of participants who used the interface at each point on the path: darker colours mean that several participants used the interface at that location, while lighter colours mean that only a few participants used the interface at that location (white means that no participant used the interface).

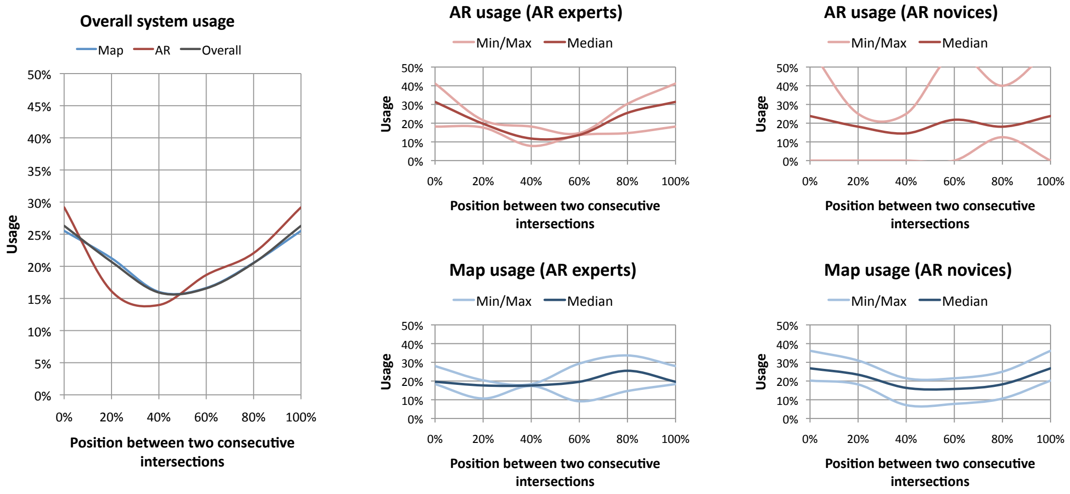
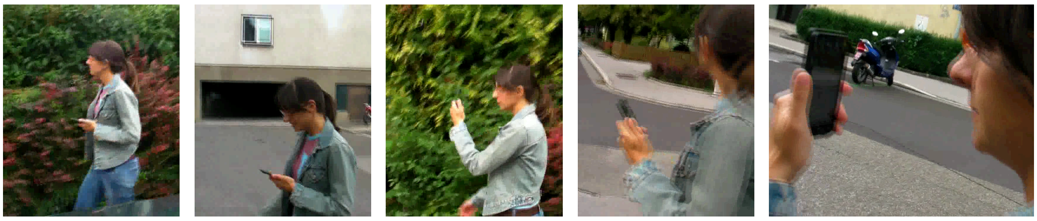


Figure 7.5: The average usage of map and AR interfaces between consecutive intersections. Left: the overall system usage. Middle: the AR and map usage for the participants with previous AR experience. Right: the AR and map usage for the participants without previous AR experience. In the plots, a position close to 0% means just after an intersection, whereas close to 100% means approaching the next intersection.



(a) (b) (c) (d) (e)

Figure 7.6: Use of the navigation system while walking through an intersection. This figure shows a typical behaviour at an intersection: (a) the participant approaches the intersection; (b) she checks the map; (c) she switches to AR; (d) she crosses the street, and turns into a side street; (e) she finally checks the AR view, once more, to verify that she turned into the right street.

experts increases more mildly near to road intersections. In contrast, the plot for AR novices shows more use of the map before and after intersections: this suggests that the map supported AR novices in proximity of intersections, similar to how AR supported AR experts. We cannot comment on the use of AR between intersection by AR novices, because the data has a high variance due to the small number of measurements. Interestingly, Figure 7.4 also shows increasing use of AR just before reaching the end of the path. We gathered informal feedback from our participants on that they would have liked an AR cue highlighting the final destination they had to reach. The final destination is therefore clearly another scenario where users would need AR support.

Usage behaviours. Both the AR experts and novices used the navigation system while walking. Only the three AR experts stopped to use the navigation system, at a total of seven different points in time (four for p1, two for p2, and one for p4). In Figure 7.7, we show a detailed timeline of how participants behaved in these cases. As can be seen, in all cases the participants started using the system while walking, and only stopped and used the system while standing still after a failed attempt to use it while walking (this behaviour is also shown in Figure 7.8). Usage while standing still is probably only plausible at difficult decision points, when the user does not succeed in making a decision on the fly while walking. In all the seven cases, the interface used for making or confirming the final decision was AR.

When participants used the system while walking, tracking relied on sensor data which

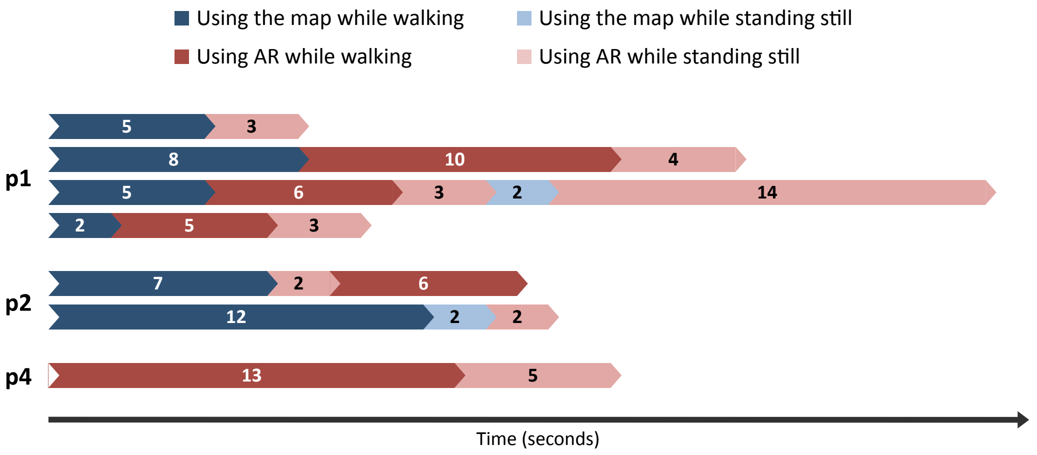


Figure 7.7: The seven cases in which participants stopped walking while using the system. In our experiment, most participants only used the system while walking. Only three participants stopped walking while using the system. This happened seven times (4 times for p1, twice for p2, and only once for p4). In this figure, we plot detailed timelines of usage behaviour for all these seven cases (all timings are in seconds).

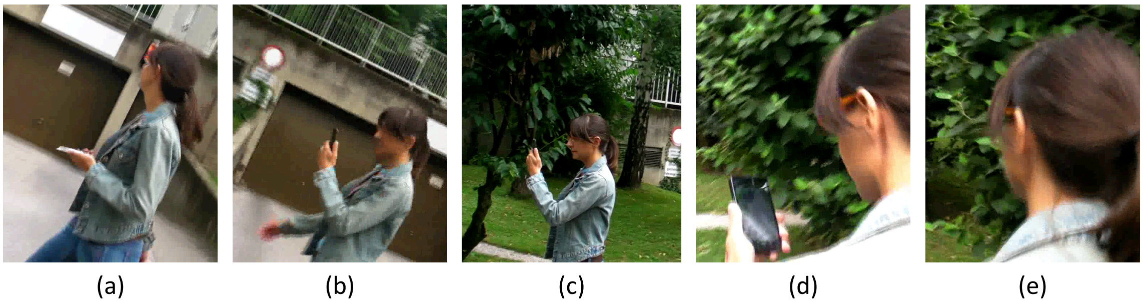


Figure 7.8: One participant stops walking while using the system. This figure shows the participant's behaviour in this case: (a) the participant is walking and not looking at the device; (b) she checks the map while walking; (c) she switches to AR, stops walking (slightly confused) and focuses her attention on the AR view; (d) she starts walking again, still looking at the AR view; (e) she finally puts the phone away, and keeps walking.

caused visualisation inaccuracies. Participants interpreted unintentional misplacements of the arrows as intentional instructions. For example, one participant (AR novice) interpreted a left-turn arrow with positional offset as an instruction to cross the street and turn left onto the opposite pavement. Two other participants (one AR expert, one AR novice) interpreted errors in the orientation of the arrow as instructions to leave the pavement and walk on the street, or to move back from the street onto the pavement. Comments from the subjects hinted that the affordance of the AR view increased participants' expectations about the accuracy of the visualised information.

7.3 Reflection on our research questions

The preliminary results presented in this chapter provide valuable understanding about the scenarios in which users need outdoor wayfinding support from handheld AR. Yet, the AR view had a low level of use in our study, due to the fact that the map interface was deemed good enough for our experimental task. From the interviews, we found that two factors are principally responsible for this low adoption: (1) the difficulty of the task and (2) the instability of the augmentations. We deduce that AR views must target navigational tasks in which common maps are challenged and at the same time provide a trustful experience, to be an enhancement over more common interfaces such as maps.

If we consider the results of this chapter from the perspective of our main research questions (Section 1.3), we can see that users access the AR view mostly when they need to make a decision, typically in proximity of road intersections or approaching their final destination. These are therefore the most important scenarios where AR can enhance outdoor navigation, and the most important locations to support with accurate tracking. It must also be considered that users seek for support both before and after an intersection, thus giving AR the dual role of providing instructions (e.g., where to turn) and confirmation (e.g., if one turned into the right street).

Our results also show that a naïve AR visualisation fails when tracking accuracy is poor, misleading users up to the point of convincing them to walk in the middle of the street, rather than on the pavement. Since the AR interface was used while walking in our experiment, we deduce that supporting a walking user with accurate tracking is fundamental. However, as continuous accurate tracking is a known hard problem, a more applicable alternative solution is to inhibit AR usage at the interface level, resorting to an interface different than AR (like a map) whenever users are walking. In general, tracking accuracy must be clearly communicated in the interface: the AR visualisation must diminish user expectations about the accuracy of the given information, in order to provide an affordance that does not go beyond the real accuracy of tracking technology.

Chapter 8

Indoor navigation

In this chapter, we discuss indoor navigation. In contrast to outdoor navigation, there is currently no global tracking solution (such as GPS) available indoors, thus navigation systems must instrument the environment with complex infrastructure to be able to continuously track the user's position. When this is possible, we reason that results for outdoor navigation systems (Chapter 7) also apply to indoor navigation systems. This chapter investigates an alternative solution to continuous localisation, *sparse localisation*, which is applicable when it is not possible to instrument the environment with complex infrastructure. We support users with detailed navigational instructions only at certain *info points* in the building, and with less-detailed information in-between. Since localisation only takes place at these info points, sparse localisation has lower instrumentation requirements and is therefore more easily applied to generic indoor environments.

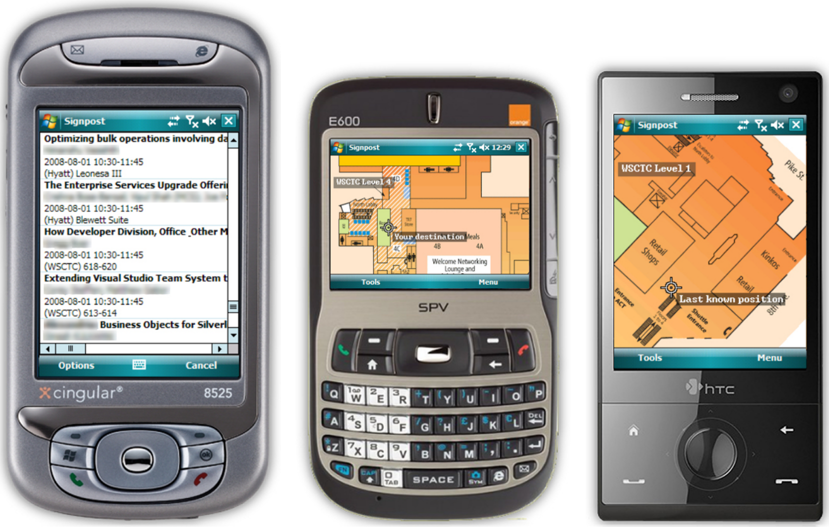
We present three design iterations of an indoor navigation system based on sparse localisation. In the first, we investigate the usability of an indoor navigation system that supports users only at sparse locations in the building. In the second, we look at how AR cues can enhance such a system. Finally, we elaborate on our previous findings and propose a design based on Mixed Reality (MR) cues, that supports users' continuous navigation despite the sparse localisation scenario. After each design iteration, we conduct experiments to explore the usability of the interface, motivating and supporting the next iteration.

8.1 Sparse localisation

We developed a location-based conference guide called *Signpost* to gain insight about the usability of an indoor navigation system based on sparse localisation. Events in large spaces, such as conferences, challenge participants to find their way through vast multistorey convention centres or hotels. The large scale of the conference venue and the need for setting up the venue efficiently typically prevent the deployment of dense localisation infrastructure. We therefore rely on localising users only at specific locations in the environment.

8.1.1 Interface design

Signpost combines a conference calendar with a navigation system (Figure 8.1). All calendar entries are linked to locations, so that users can plan their fastest route from the current location to the desired lecture hall. The results are displayed on a map that can be freely panned, rotated and zoomed.



Optimizing bulk operations involving da
[redacted]
2008-08-01 10:30-11:45
(Hyatt) Leonesa III

The Enterprise Services Upgrade Offeri
[redacted]
2008-08-01 10:30-11:45
(Hyatt) Blewett Suite

How Developer Division, Office _Other M
[redacted]
2008-08-01 10:30-11:45
(WSCTC) 618-620

Extending Visual Studio Team System t
[redacted]
2008-08-01 10:30-11:45
(WSCTC) 613-614

[redacted] Business Objects for Silverl
[redacted]

(a) (b) (c)

Figure 8.1: Examples of phones running Signpost and screenshots of the application. We show typical views of the application: (a) conference schedule showing the location of each talk, (b) map showing the destination of the user, based on a chosen talk, and (c) map showing the last known position of the user in the building.

In Signpost, we use marker-based localisation: when a user points the camera of her device at one marker (see Figure 8.2, right), Signpost localises her in the building and updates her position in the map view. We implemented this feature of Signpost using our marker tracking library [199], running on Windows Mobile devices.

Installing Signpost at a new venue requires creating one or more 2D maps of the venue, and a database of marker locations and orientations on the maps. The most efficient way to implement sparse localisation is to glue the markers onto poster stands, which can be quickly deployed on-site at the planned locations (Figure 8.2). A coarse deployment of the

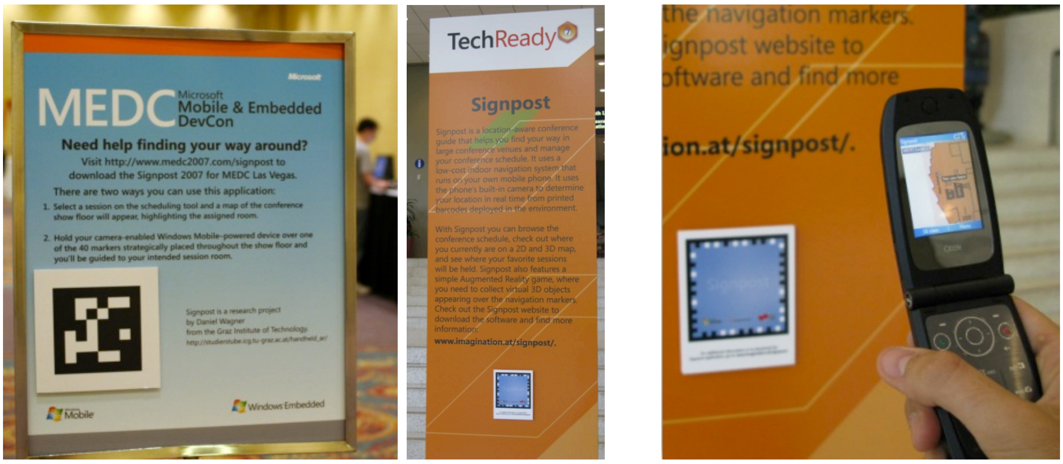

Microsoft
MEDC Microsoft Mobile & Embedded DevCon
Need help finding your way around?
Visit http://www.medc2007.com/signpost to download the Signpost 2007 for MEDC Las Vegas.
There are two ways you can use this application:
1. Select a session on the scheduling tool and a map of the conference show floor will appear, highlighting the assigned room.
2. Hold your camera-enabled Windows Mobile-powered device over one of the 40 markers strategically placed throughout the show floor and you'll be guided to your intended session room.
Signpost is a research project by Daniel Wagner from the Graz Institute of Technology http://studierstube.icg.tu-graz.ac.at/handheld_ar/
Mobile
Windows Embedded
TechReady
Signpost
Signpost is a location-aware conference guide that helps you find your way in large conference venues and manage your conference schedule. It uses a low-cost indoor navigation system that runs on your own mobile phone. It uses the phone's built-in camera to determine your location in real time from printed barcodes deployed in the environment.
With Signpost you can browse the conference schedule, check out where you currently are on a 2D and 3D map, and see where your favorite sessions will be held. Signpost also features a simple Augmented Reality game, where you need to collect virtual 3D objects appearing over the navigation markers. Check out the Signpost website to download the software and find more information:
www.imagination.at/signpost/.
the navigation markers.
ignpost website to
oftware and find more
ion.at/signpost/.

Figure 8.2: Marker-based localisation with Signpost. Left: posters and fiducial markers deployed at the MEDC 2007 and TechReady7 conferences. Right: our localisation system at TechReady7: a user points the phone towards one marker, until the marker is recognised and the system updates the map highlighting the user's position.

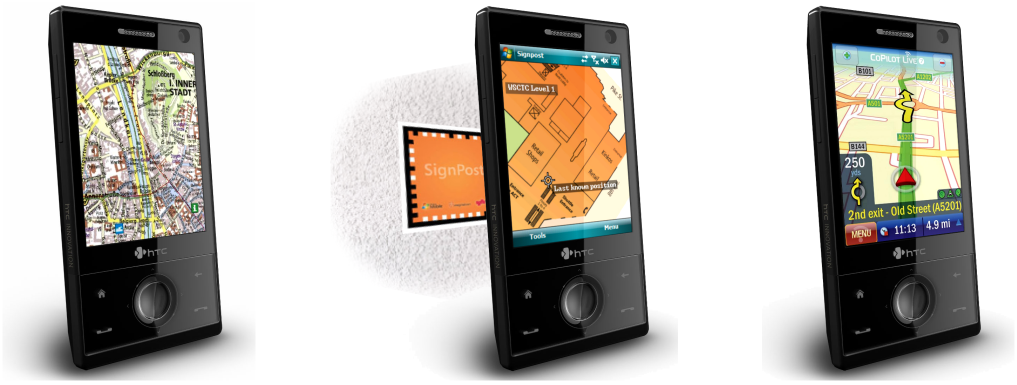
poster stands (±50cm) is usually sufficient, as small displacements will not be noticeable on the map. Sparse localisation therefore limits the required infrastructure to a small number of poster stands in the buildings. For example, only 37 poster stands were needed to set up Signpost in an area of roughly 100m × 200m, at a conference in the Venetian Hotel, Las Vegas.

In contrast to a regular conference guide, our design gives users a chance to retrieve live positioning information at special info points in the environment. In comparison to a system using continuous localisation, though, Signpost cannot provide live position information when users are not at an info point.

8.1.2 Controlled study

Before deploying Signpost in a real-world scenario, we first conducted a controlled study to evaluate the usefulness of our sparse-localisation approach in comparison to a system with no localisation, and a system with continuous localisation. These conditions represent the extremes of a localisation continuum, as shown in Figure 8.3, whereas sparse localisation is located in-between the two.

Task. Participants were asked to use the digital map and the localisation system as their only aid to reach a specific pre-defined destination. The location of the study was our university department, a complex composed of four three-storey buildings connected by several bridges. It contains many repeated features with a general lack of clear landmarks.



No localisation
(e.g. map browsers)

Sparse localisation
(e.g. Signpost)

Continuous localisation
(e.g. GPS-based systems)

Figure 8.3: Continuum of localisation techniques. We position sparse localisation on an ideal continuum, that spans between systems with no localisation and systems with continuous localisation.

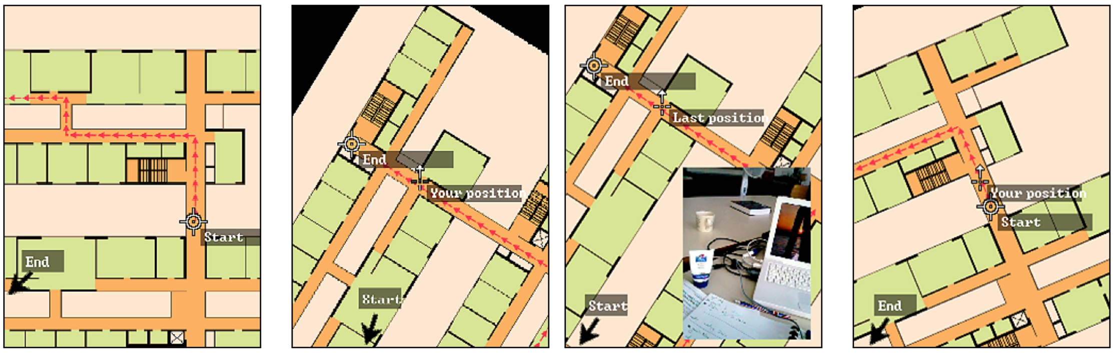
We consider it to be a significant example of a "hard case" for navigation in a new environment. We selected three different destinations, balancing their difficulty in terms of distance from the start point, bridges to cross and number of in-between floors.

Independent variable. We had three conditions for this experiment (Figure 8.4):

- *No localisation.* We implemented a map view that can be panned with a finger on the touchscreen.

- *Sparse localisation.* We integrated our marker-based localisation solution into the map view and presented the live camera video view in a screen corner (see Figure 8.4 (b)). The position of all info points are presented on the map as red dots. As soon as a marker is detected, we automatically update the position and orientation of the user, presenting it as a labeled icon on the map. However, we do not re-position and re-orient the map automatically – a pilot study revealed that users prefer to rotate and centre the map manually when using a sparse localisation system, because a sudden re-orientation of the map causes a loss of spatial awareness.

- *Continuous localisation.* We used a Wizard of Oz approach, as we did not have an indoor equivalent to GPS. One examiner walked behind the participant and remotely controlled the participant's phone, continuously updating the position and orientation of the icon on the map (Figure 8.4 (c)) to match the position and orientation of the participant in the building. The device was remotely controlled via a second device, connected to the participant's device using Bluetooth. Continuous



(a) (b) (c)

Figure 8.4: Screenshots of the application used in the controlled study. We show screenshots of the three conditions: (a) no localisation, (b) sparse localisation and live-video view for pointing the camera at the marker, and (c) continuous localisation.

> localisation is a control condition only useful for comparison within our experiment, and the Wizard of Oz approach allowed us to quickly build a running system.

In all localisation modes, the map shows the start and destination points with crosshairs. When such locations are outside the view of the map, we present off-screen directions using labeled arrows. Users can access the map of a specific floor by pressing the number key corresponding to the desired floor number.

Hypotheses. We hypothesised that continuous localisation is found easier to use than the other systems and provides the highest user confidence because, in contrast to the other two conditions, it provides continuous interactive feedback on the user's position in the environment. We also hypothesised that sparse localisation provides higher user confidence than no localisation because, in contrast to that condition, it provides interactive feedback on the user's position in the environment at least at a few locations in the environment. However, we expected sparse localisation to be found to require more learning effort, because of the need to learn how to operate marker tracking. The basic hypothesis of this experiment is that sparse localisation is a better solution than no localisation, if continuous localisation is not possible.

Procedure. We used a within-subjects design: all participants navigated to all three destinations, each time with a different localisation mode. We counterbalanced the order of localisation modes and destinations using a Latin square, to avoid biases. We gave all users some time to familiarise themselves with the application before starting the evaluation.

After completing all the three tasks, we asked users to rank the three conditions from

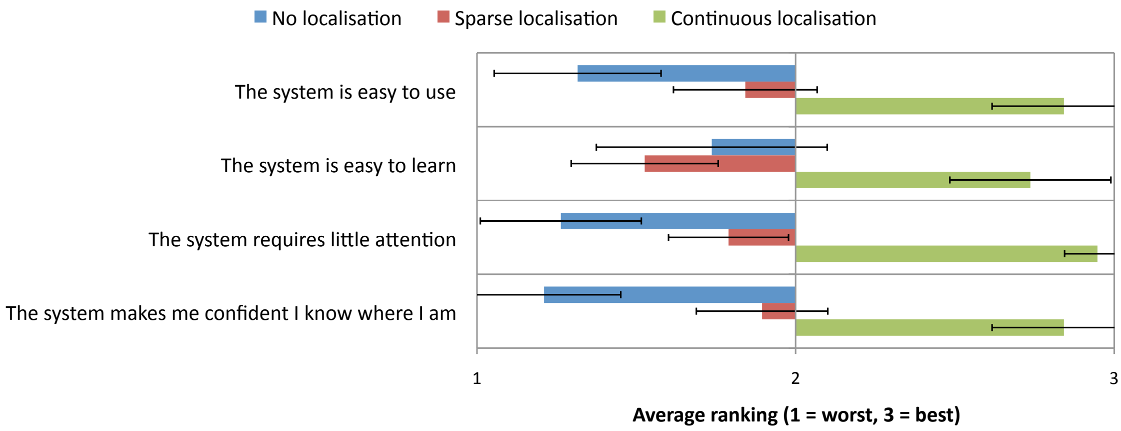


Figure 8.5: Subjective rankings of the three localisation systems. The three systems were ranked on four different criteria.

worst to best. The conditions were ranked according to four different criteria: ease of use, ease of learning, required attention and confidence about current location.

Participants. We recruited 20 users (10 male and 10 female) with diverse cultural background, varying expertise in technology, and aged between 20 and 34 years old (average of 25). In order to avoid biasing of the results, we ensured that no user had previously been inside the buildings.

Results. For each criteria and participant, we assigned a score of 1 to the worst condition and a score of 3 to the best condition. The average ranks and their 95% confidence intervals are presented in Figure 8.5. A Friedman's test shows that the effect of the localisation mode is significant on all criteria ($p < .001$).

A post-hoc analysis with Bonferroni correction shows that almost all pair-wise differences are significant. Continuous localisation was found to be easier to use and to learn, to require less attention, and to provide more confidence than the other two localisation modes ($p < .001$, for all differences). Sparse localisation was found to be easier to use than no localisation ($p <.05$), to require less attention ($p <.01$) and to provide more confidence ($p < .001$), but not easier or harder to learn.

There was no significant difference in task completion time with the three conditions (repeated-measures ANOVA, $F_{2,36} = 2.34$, $p = .11$), because of the high variation in individual results. Participants took on average 5.32 minutes with no localisation ($SD = 3.48$), 4.26 minutes with sparse localisation ($SD = 2.13$), and 3.47 minutes with continuous localisation ($SD = 2.17$).

Discussion. The results confirm our hypothesis on continuous localisation, showing that it is found easier to use and it provides higher confidence than the other two conditions. As hypothesised, the results also show that user confidence is significantly higher when using sparse localisation, compared to no localisation. In contrast to what we hypothesised, our results show no significant difference in the ease of learning between the two conditions – this suggests that operating the marker-based localisation has no impact on the perceived difficulty of learning to use the interface.

The results also show that users find sparse localisation significantly easier to use than no localisation, and to require less attention. During the experiment, we noticed that the information provided at sparse localisation points helps users mentally registering the view on the digital map with the real environment. While users did not use sparse localisation intensively when they were going in the right direction, it seemed fundamental for users who were lost, in order to re-map their mental model with the real building and re-structure their path accordingly. While with no localisation users had to match landmarks in the environment with landmarks on the map, with sparse localisation the burden was reduced to registering the icon on the map with their real position and orientation in the environment. We conclude that compared to a static map, users feel sparse localisation to be easier to use as it provides a quick means of verifying their position, even if only at sparse locations in the building.

8.1.3 Exploratory study

Our previous results confirm the validity of our sparse-localisation approach, but they do not give us insight on how it would apply to a real-world large-scale scenario. We therefore conducted one further exploratory study in larger environments, deploying Signpost at a number of international conferences: MEDC2007 (April 2007), TechEd2007 (July 2007), TechReady6 (February 2008) and TechReady7 (July 2008). Over the four conferences, more than thousand distinct users installed Signpost on their devices. This gave us the chance of collecting feedback from a large number of real users in a natural environment via usage logs, questionnaires, on-field observations and interviews. We interviewed a limited number of users personally, and collected questionnaires and usage logs from a larger part of them. Our overall aim was to find out how useful attendees found Signpost, but for the purpose of this thesis we will focus in particular on the acceptance, the level of adoption and the perceived usefulness of our sparse localisation approach.

At MEDC, we collected 34 anonymous questionnaires. In the questionnaires, we asked

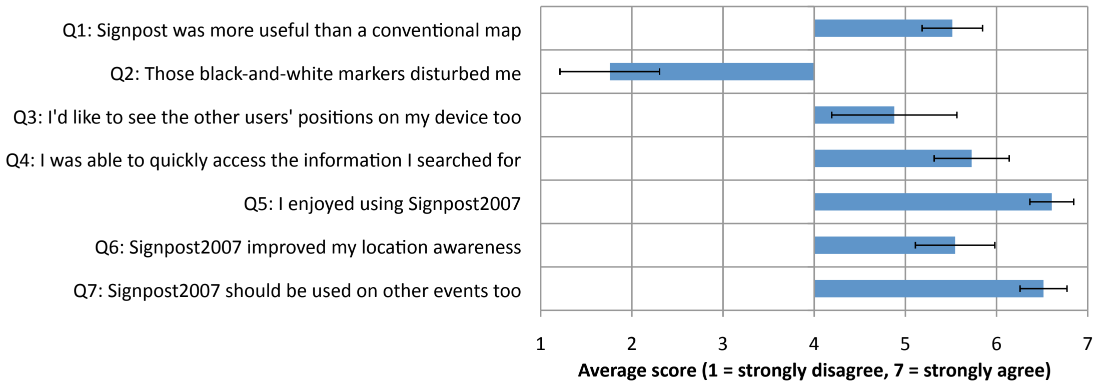


Figure 8.6: Subjective questionnaire ratings from the MEDC conference. Questions touched both the usefulness of Signpost and the acceptability of sparse localisation targets.

attendees to answer a number of questions on a 7-point Likert scale ranging from 1 (“strongly disagree”) to 7 (“strongly agree”). The results are presented in Figure 8.6, while the questionnaire is presented in Appendix A.5. The results show us that users found Signpost more useful than the conventional conference map (Q1), which was part of the printed conference booklet. Furthermore, users also felt an improvement in their location awareness (Q6). Using markers in public areas could raise questions concerning visual clutter, but users were mostly not disturbed by them (Q2). All other questions also received positive answers. Overall, these results give us a first indication of the good acceptance of the markers and of the perceived usefulness of sparse localisation, with respect to a traditional conference map.

At TechReady6, we could also conduct semi-structured interviews with a number of conference attendees. One part of the interview focused on how well the navigation worked. While the marker tracking system is accurate, the markers themselves were mounted coarsely to keep the effort of mounting and measuring to a minimum. Yet, users were generally satisfied with the tracking accuracy: “when I looked at it, immediately I thought wow, this is where I am.” One user said that tracking “was accurate enough. [...] Two feet off the door versus four feet off the door really doesn’t matter.” Attendees also suggested adding step-by-step instructions: “I think the biggest thing that would help me was if it would tell me steps: go down escalator, turn right, ... like some of the car navigations things, but maybe not that precise”. Overall, attendees liked the way Signpost supported their navigation and were satisfied with the accuracy of the system, but they would have appreciated receiving coarse navigational instructions also in-between localisation points.

At TechReady7, we collected software logs from 74 anonymous users over four days,

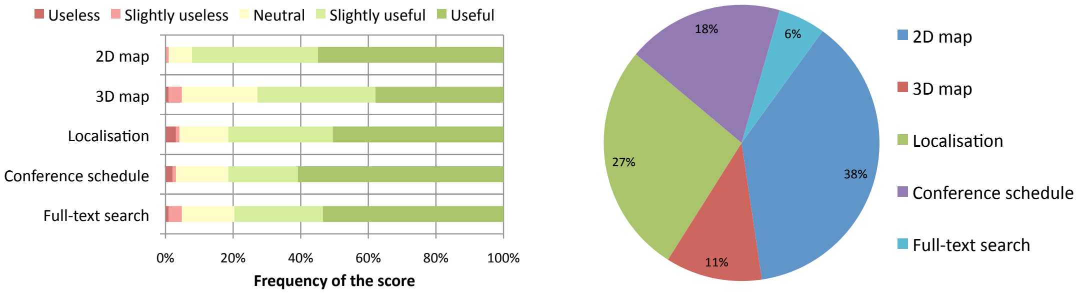


Figure 8.7: Results from the data collected at the TechReady7 conference. Left: subjective user ratings for the usefulness of each function. Right: usage statistics for the five functions.

to obtain quantitative understanding on how often the different functions of Signpost are used. For this purpose, we identified five core functions: 2D map, 3D map, localisation, conference schedule and full-text search (of the schedule). We looked at the number of times the functions were invoked, and triangulated the log data with questionnaires (see Appendix A.6), collecting the perceived usefulness of the functions from 64 distinct users. In the questionnaire, we asked users to rate the usefulness of the five functions, in comparison with the printed conference booklet, on a 5-point Likert scale ranging from 1 ("useless") to 5 ("useful"). The results (Figure 8.7) show that 2D map and localisation are the most triggered functions. From the questionnaires, we also see generally high rates for the perceived usefulness of all functions, as compared to the printed conference booklet.

8.1.4 Discussion

The results from our controlled study show that an indoor navigation system that uses sparse localisation is found to be easier to use and makes users more confident than manually operated floor plans. Sparse localisation is therefore a valid alternative when continuous localisation is not possible. The deployment of Signpost at a number of large-scale conferences not only confirms the validity of sparse localisation, but also its applicability in real-world scenarios. To our knowledge, Signpost is the first indoor navigation system successfully deployed at several large-scale venues and installed by more than one thousand users on their own device.

8.2 Sparse localisation and Augmented-Reality cues

Our first results highlight the value of sparse localisation as a support for navigation, when continuous localisation is not viable. However, Signpost only gives simple localisation information when a marker is visible by the camera of the device. When navigating between one marker and the following one, users receive no instructions from the system. The lack of navigational instructions also emerged during the interviews, and was seen by users as a shortcoming of the system. As the focus of this thesis is on how AR can enhance navigation, a natural next step for our research work was to investigate how AR can be used to continuously provide navigational instructions in a sparse-localisation system.

8.2.1 Interface design

We set three key requirements to inform the design process of our interface:

- Minimal instrumentation of the environment (sparse localisation).
- Continuous navigational support as a sequence of turn-by-turn instructions.
- AR navigational cues adaptive to localisation accuracy and user activity.

This section presents the choices that led to the proposed design, shown in Figure 8.8. We recruited experts and external users for a pilot task to inform and refine the design of our prototype. The pilot task consisted of collecting a box from an office room and delivering it to another nearby office room, and required a number of turns and one floor change over a distance of about 100 steps. We detail the space of design possibilities and we explain why we chose one over the other, for each of the three key requirements.

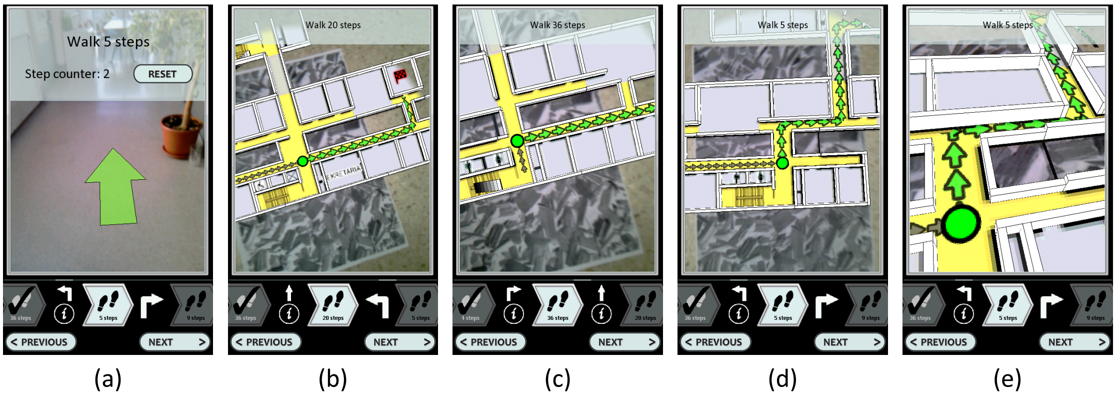


Figure 8.8: Screenshots of our prototype. (a) As a user is walking, the interface presents sketchy information on the current activity and directional information as a perspective arrow. (b-e) As a user stops at an info point, AR information is presented as a World-in-Miniature (WIM). From afar (b-d), the WIM appears as a 2D map. In (b), the target office room is visible and marked with a red flag. From closer and tilted perspectives (e) it is possible to examine the WIM in 3D.

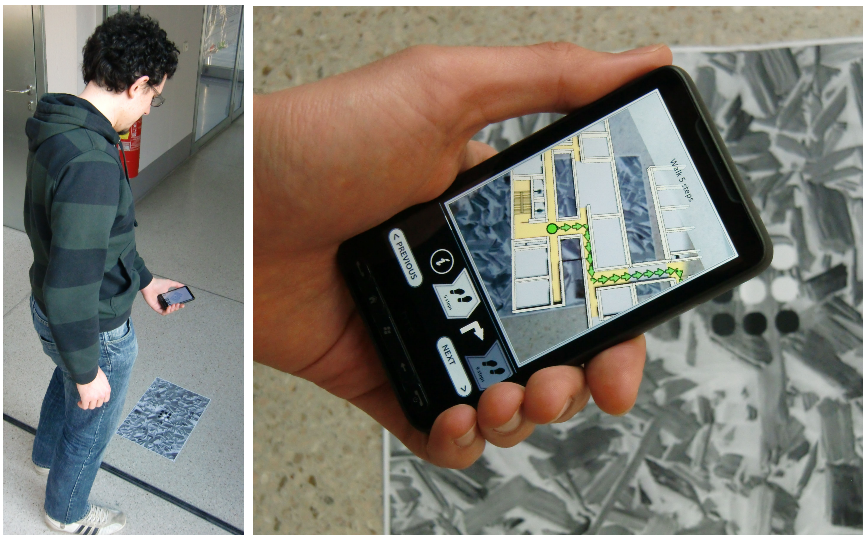


Figure 8.9: Info points and their placement on the floor of the building. (Left) User accessing an info point. (Right) Corresponding view on the phone of the user.

Info points. Just like the previous design of Signpost, we use sparse localisation at selected info points in the building, although we mounted the tracking targets on the floor, rather than on walls. During the pilot, we found that floor-mounted artificial markers are not sufficiently robust against lighting changes (e.g., reflections of the neon lights in the building) and partial occlusions (e.g., the foot of the user partially covering the target). In our new design, we therefore use natural-feature tracking targets (Figure 8.9), which are more robust to these types of common issues.

As observed in our previous experiments, the info points act as spots where users can ensure they are on track. Similarly, some pilot participants went straight to the nearest info point to reorient their interface, when they were unsure about the next steps. We noticed that it is important to provide info points in proximity to decision points. Hence, the density of info points is dictated by the building layout and the occurrence of decisions points. In the case of our buildings, those decision points are usually near staircases or where multiple paths propagate.

The proposed approach of floor-mounted info points relies on a sufficiently distinguishable floor texture, thus info points can be carefully designed to be part of the overall building design. Furthermore, the info-point identifier can be made completely transparent to the user by means of wireless technology. In Figure 8.10, we present a conceptual design that fits a modern shopping-centre scenario.

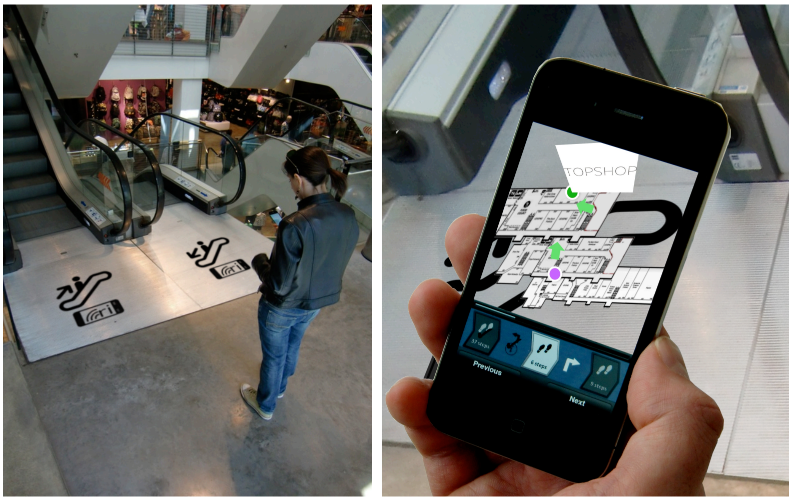


Figure 8.10: Conceptual design of an info point for a shopping center. Left: a user accessing the info point. Right: corresponding view on the phone of the user.

Turn-by-turn instructions. We exploit basic human navigation abilities to provide continuous support for navigation in our sparse-localisation scenario. People are used to follow linear sequences of turn-by-turn directional instructions, if sufficient context information is provided (in particular, environmental features pertaining to the decision points on the path [2]). In a recent paper, Brush et al. [18] describe a navigation system that supports users only with a static list of instructions to be performed – e.g., walking a number of steps in a certain direction, or going up a number of floors. The results of their evaluation show that this is a viable solution. We therefore chose to combine info points with a list of turn-by-turn instructions, exploiting human abilities over short paths to compensate for the lack of localisation in our system.

The *instruction view* (Figure 8.8, lower part of the screenshots) presents a sequential list of instructions that a user needs to perform to reach the target destination. Supported activities are *walk*, *change floor*, *turn* and *reach office*. Instructions are shown as a sequence of arrow-shaped elements, pointing left-to-right to communicate their sequential ordering. We visualise turns and info points between the instructions and we clearly identify with a checkmark which instructions have already been performed. Users can scroll through the instructions using the touch-screen of the device. Users can also switch to the next or the previous instruction with a single tap – either on a button (on the bottom part of the instruction view) or on the instruction itself. On the upper part of the instruction view, a small progress bar indicates the current progress in the navigational task.

Figure 8.11 shows all the consecutive design iterations of the instruction view. After the

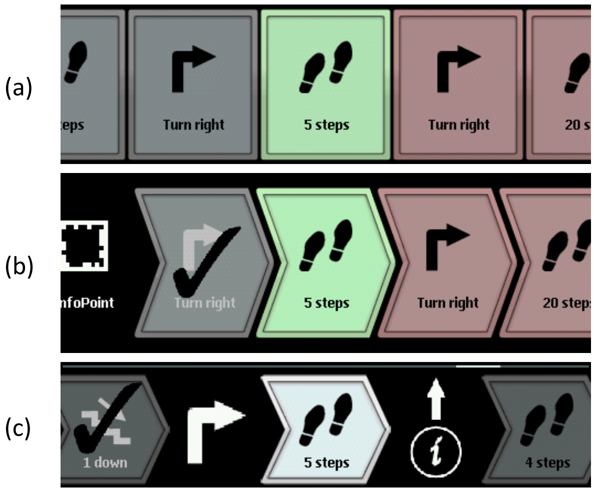


Figure 8.11: Iterative designs of the instruction view. We show the first (a), second (b) and last (c) design. In (b-c) we clearly convey the flow of instructions from left to right. We also add the position of the info points between instructions. In (c) we do not show turns as instructions, to eradicate the need for a double tap at each turning point.

first pilot study, we adapted its design based on two issues we observed. First, we observed that all instructions were perceived as equally relevant, hence we tried to communicate the flow from left to right more clearly, to distinguish the already performed instructions (left) from the current one (middle, highlighted) and the ones to be performed next (right). We support the idea of flow by changing the design of our instruction buttons, shaping them as arrows that point towards the instructions to be performed next. We use checkmark icons to clearly label already performed instructions. Secondly, we add the info points to the list of instructions. During the pilot study, we noticed that users had problems tracing back the instructions they had performed after departing the last info point. Adding the info points to the instructions helps users to trace back what they did after the latest "safe position". After a further pilot study, we decided not to represent turns as instructions, as this requires an unnecessary second tap on the next button at each turning point: first to advance to the turning instruction and then to advance to the following walking instruction.

In a preliminary version of this design, we aimed at automatically detecting if the user had completed an instruction, using information from a step counter and from the phone's sensors. Our objective was to let the list of instructions flow automatically as the user completed consecutive instruction. As a fallback, users could manually advance or return to an instruction. After the first pilot, we removed this automatism from our design, as users found any automatic switch confusing when triggered erroneously by unpredicted causes (e.g., longer or shorter stride lengths, magnetic influences on the phone's digital compass). We reason that the adopted one-tap solution to advance instructions is a valid compromise between the feedback given by the interface and minimal requirements of manual user interaction.

AR cues. The *AR view* (Figure 8.8, upper part of the screenshots) augments the environment with AR cues regarding the instructions to be performed. We adapt the visualisation depending on the localisation accuracy and the type of activity the user is engaged in (either walking or standing still).

Whenever the system detects that the user is walking, we automatically enlarge the textual information in the AR view (Figure 8.8 (a)). Walking users must divide their attention between the physical movement and the use of the interface, and we avoid overloading them with information. We therefore provide easy-to-read, sketchy information that details exactly what the user must do in order to perform the current instruction. We also embed an indicative step counter in the view for the user's convenience. In this case, the AR view provides directional information in an egocentric frame of reference: we overlay the video with a perspective arrow always visible in the centre of the view, spatially oriented to lie flat on the ground and to point towards the current walking direction. Its purpose is to give egocentric feedback on the walking direction, when the system does not have knowledge of the location of the user. This visualisation is analogous to floor signs commonly found in public areas to direct people's navigation.

Whenever the system detects that the user stops over an info point, we shrink the textual information to make more room for the AR visualisation (Figure 8.8 (b-e)). As the user is not walking while accessing the info point, we assume that more attentional resources are available and we provide more detailed and complex information using a WIM. A novel aspect of our approach is that tracking targets are installed on the floor of the building, rather than on the wall as in previous work (e.g., in Signpost or in the work by Müller et al. [126]). In this way, a WIM augmented on the tracking target appears aligned with the building (see Figure 8.9) and no mental rotation is required by the user. We overlay the WIM with dynamic information about the location of the user (green circle), the path that the user must follow (in green) and the direction from which the user approached the info point (smaller, in grey). Finally, if the target office room is near the info point, we also highlight it (red flag). We designed the WIM by extruding walls and stairs from a 2D map, so that it appears as a 2D map from afar but if needed it can be explored in 3D from closer and tilted points of view (see Figure 8.8 (e)). The rationale is that a 2D map can quickly convey route information, whereas a WIM supports landmark recognition, as shown by Kray et al. [86]. When seen from above, the floor texture of the WIM provides further details, such as toilets, wall shadows and the location of doors. By showing the WIM in AR, we provide an easy affordance of browsing the path using the mobile phone, panning the device over the tracking target to control the viewpoint on the WIM, similar to our work on augmented maps (Chapter 6). In contrast, exploring a VR model in this detail would require a 6-DOF manual control of the viewpoint over the WIM, which would be much more cumbersome.

8.2.2 Controlled study

We conducted a user study to validate our interface, to explore how people use it, and to study how the presence of augmented info points affects the performance of users, in comparison to a static list of turn-by-turn instructions.

Task. Similarly to the controlled study in Section 8.1.2, we asked participants to find one office room inside our department buildings. We reason that this is a typical scenario for users who must navigate an unknown indoor environment, e.g., an office building or a hospital. Our department is composed of four buildings that are interconnected to each other by several bridges. All buildings have a strong cubature with in-situ concrete walls and internal patios, and they contain a large number of repetitive features with a general lack of clear landmarks. There is virtually no signage for departments or offices. We consider it a hard case for indoor navigation.

Independent variable. We had two conditions for this experiment:

- *No info points (**NoIP**).* The condition in which info points are absent is a baseline condition, analogous to the work presented by Brush et al. [18]. In this condition, the system provides users with a sequence of turn-by-turn instructions that must be followed to reach the target office room. There are no info points accessible throughout the path.

- *With info points (**IP**).* The condition in which info points are present and the system provides users with the same sequence of turn-by-turn instructions as in NoIP. In this condition the sequence of instructions includes information on the available info points through the path. During the study, we physically placed a number of info points in the building. Users could move to a nearby info point to re-orient

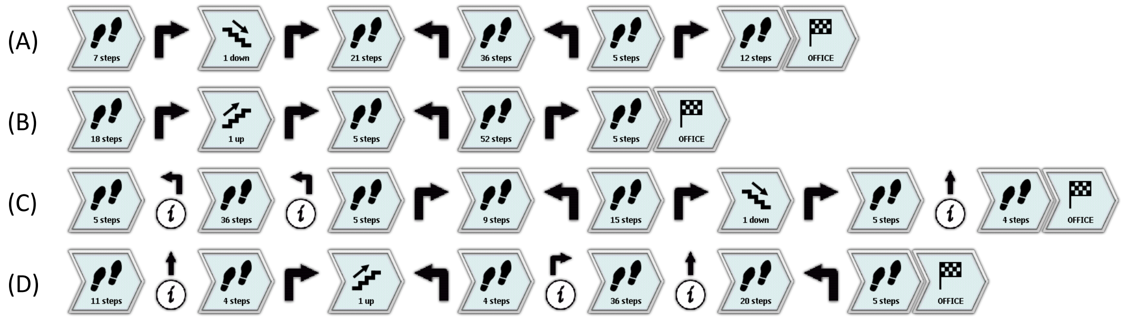


Figure 8.12: The tasks designed for the user study. Tasks A and B were designed for the NoIP condition, tasks C and D for the IP condition. All tasks have comparable length and difficulty. IP tasks appear longer as straight path segments are split in two by info points.

the interface and themselves, obtaining detailed location-aware information on the upcoming path segments.

Hypotheses. We expected the presence of info points to keep users more on track and to lower their perceived workload, as also observed in our previous studies. We therefore set the following hypotheses: (H1) shorter walked distances in IP compared to NoIP, (H2) lower number of navigation errors in IP compared to NoIP, and (H3) lower workload perceived in IP than in NoIP.

Procedure. We used a within-subjects design: all participants experienced both NoIP and IP conditions. We designed 4 tasks, 2 for each condition (see Figure 8.12). All tasks have comparable difficulty, containing a similar number of turns and always one floor change up or down. We recorded a ground-truth step count for each task using a commercial step counter, and verified that all task have a similar path length (127 ± 3 steps). Participants were asked to perform all four tasks; we used a Latin square to balance the order of tasks between them. One study session lasted on average 45 minutes.

We began a study session by collecting demographic data from the participants and having them sign an informed consent form that introduced the procedure of the user study. We then had participants conduct a tutorial task that forced them to try all functionalities of the application, supported by a verbal explanation by the examiner. In particular – as AR is not a commonplace interface metaphor – we enforced AR training on two info points during the test task. We therefore assume that all our participants were at least familiar with the operation of the AR interface before starting the subsequent tasks. Finally, we conducted participants to one of the four locations designated as a starting point for the tasks, we gave them the device and we asked them to find the target office room. This procedure was repeated for all four tasks.

As the goal was to validate our design, explore how participants use it, and measure the impact of info points on task performance, we used a number of methods to collect both qualitative and quantitative data. On the device, we ran a software logger continuously recording the application status and selected events. We asked the participant to wear a commercial step counter and we recorded the number of steps needed to complete each task, for comparison against the ground truth. An evaluator followed the participant and noted all observations and all spontaneous feedback given by the participant while performing the task. The evaluator also noted on a map all occurrences of navigation errors. After each task, we asked the participant to fill in a one-page NASA TLX questionnaire [51].

After all office rooms were found, we interviewed the participants. First, we asked for subjective feedback – whether having or not having the info points changed anything in their navigation experience. The question was aimed at collecting feedback on whether users found info points useful, and if so why. We structured the rest of the interview around all noted occurrences of navigation errors, to collect subjective comments on the issues that occurred in each situation.

Apparatus. We implemented our interface in an application running on an HTC HD2 smartphone at interactive frame rates (20–30 frames per second). The integrated sensors of the phone (an accelerometer and a magnetometer) assist in estimating the device's orientation and counting the user's steps. The device's orientation is estimated using the gravity vector measured by the accelerometer and the north vector measured by the magnetometer. We use a linear Kalman filter to reduce jitter in the sensors' measurements. We use the local magnitude variations of the accelerometer measurements to count the user's steps, as proposed by Jimenez et al. [70].

At each info point, we placed a poster on the floor containing a pattern that can be detected and tracked using computer vision technology. We track the position and orientation of the device with respect to the posters using the natural-feature tracking approach described in Chapter 3. This approach allows us to track the position and orientation of the device accurately, even in the case of reflections on the poster, or when the poster is only partially visible. In the centre of the poster, we encode a unique ID for the info point as a 9-bit BCH code (4 redundancy bits and 5 data bits). We supply the application with a graph of the corridors and office rooms in our department's buildings. A module of our application uses this graph to calculate the path between any

	Step difference	Time (seconds)	Navigation errors		
			Soft	Hard	Total
NoIP	29.75	135	19	3	22
IP	-2.25	142	9	1	10

Table 8.1: Task performance per condition. We show the median difference in step count from a pre-recorded ground truth, the median task completion time, and the total number of navigation errors.

pair of connected locations using the Dijkstra algorithm. This module can dynamically recalculate the path to the target destination whenever the user reaches any arbitrary info point in the buildings.

Participants. We recruited 10 participants (5 male and 5 female) aged between 24 and 35 (median 28) through the newsgroups of two local universities. Our participants were predominantly early adopters. All participants were not familiar with the buildings where our study took place and they had not previously been involved in our research work. We compensated all participants for their time with a voucher for a local media store.

Results. We focus our analysis on verifying our hypotheses on task performance and perceived workload. We integrate the quantitative analysis with a qualitative discussion of how participants used the interface, the problems they experienced, and how info points impacted on their navigation.

Task performance. All users completed all tasks successfully. As a metric for task performance, we looked at the step difference from a recorded ground truth, at the task completion time, and at navigation errors made by the users (see Table 8.1). Step count and task completion time are of course a rough measure of the performance in a navigation task, as they also depend on walking speed and stride length. However, our analysis only considers relative differences between conditions, therefore the measurements can be used to highlight whether each participant performed better in one of the two conditions. We encode navigation errors into two groups, soft and hard errors, based on the severity of the deviation from the path. Soft errors denote when the participant departed the path indicated by the interface, noticed the mistake, and recovered from the position in which she had departed from the path. Hard errors denote when the participant could not recover from the position where she departed from the path, but she had to roll back a number of instructions and repeat them from scratch.

The NoIP condition has a median step difference of 29.75 steps more than the pre-recorded ground truth, about 23% of the whole path length. A t-test shows that the

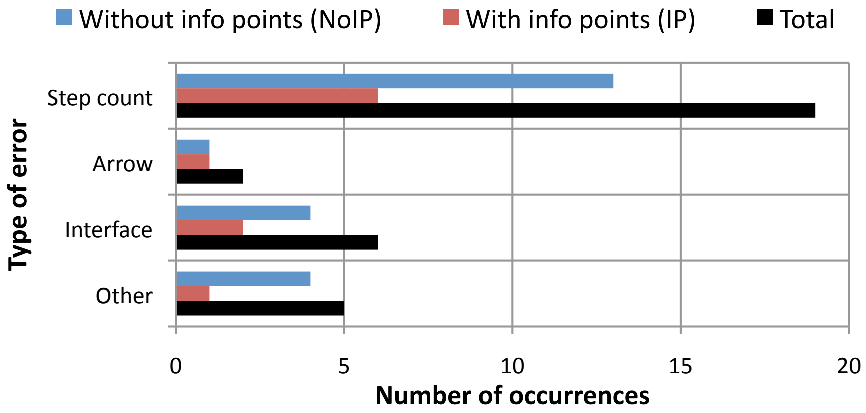


Figure 8.13: Number of navigation errors per condition and type.

difference between IP and NoIP is statistically significant (t(9) = −3.14, p = .01) and supports our hypothesis (H1). This result points to the fact that users in the IP condition followed on average an almost optimal route, whereas in the NoIP condition they usually deviated by several steps from the optimal route. While deviating more from the optimal route, the results show that users in the NoIP condition were slightly faster. This is not surprising, as participants spent some extra time browsing the augmentations at the info points, in the IP condition. A t-test shows that the difference in task completion time is not statistically significant (t(9) = −.25, p = .80).

We recorded more errors in the NoIP condition than in the IP condition. A Wilcoxon test shows that the difference is not statistically significant (Z = −1.796, p = .07). Our second hypothesis (H2) is not supported statistically, but there is a clear trend in favour of it.

From the interviews, we identify a number of error sources (Figure 8.13). We list them below, indicating the number of occurrences for each condition as (NoIP, IP):

- *Overshooting due to a wrong step count (13, 6).* Inaccuracies in the step counter and errors in the mental count of the steps caused the problem of users overshooting a turning point and erroneously going straight, in some cases performing the turn at the next intersection.

- *Confusion caused by the arrow (1, 1).* Magnetic influences within the building caused the overlay arrow to point in an incorrect direction.

- *Issues with the design of the interface (4, 2).* In four cases, participants performed the wrong activity, either because they forgot to switch to the next instruction, or because they clicked more than once on the next button. One participant confused one info point with another, and took the wrong turn. This is because info points are indistinguishable from each other in the activity bar, as shown in Figure 8.14 (a). Another participant was confused by arrows on both sides of the instruction, as shown in Figure 8.14 (b). While the flow of instructions suggests a right turn in this case, we reason that the arrows, like the instructions, should also be greyed out once the turn has been performed. Arrows in the interface have a high affordance signal and therefore easily override relevance of other parts in the interface.

- *Other (4, 1).* One participant erroneously read “one floor down” in place of “one floor up”. Another participant was distracted and missed one turn. Some other issues could not be explained or remembered by participants.

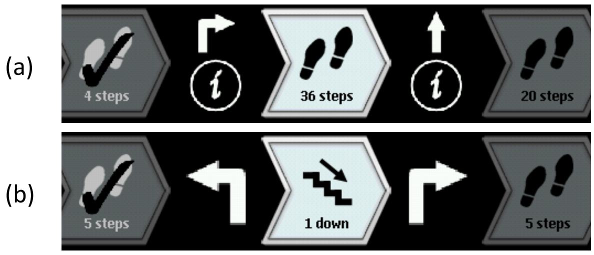


Figure 8.14: Interface ambiguities that caused navigation errors.

We investigated how users exploit the interface to cope with navigation errors, both from observations during the experiment and from interviews after the experiment. We observed that after a navigation error, participants try to match distinguishable building elements (i.e., landmarks) with the visualisation on their device, in order to recover from the error. In particular, stairs act as prominent landmarks, because they also appear as a clear checkpoint in the list of instructions. For example, users walking past an intersection with the stairs, when the next instruction was a floor change, were usually able to rapidly recover from the navigation error and go back to the stairs. In the interviews, one participant said: “when I saw that the next activity was to go up the stairs, it was obvious to me that I had walked a few steps too many.” Stairs are also a checkpoint for rolling back the list of instructions in the case of hard errors. For example, one participant was disoriented in the NoIP condition, and rolled back instructions to the last floor change, went back to the last set of stairs and then once again performed all instructions from that point on.

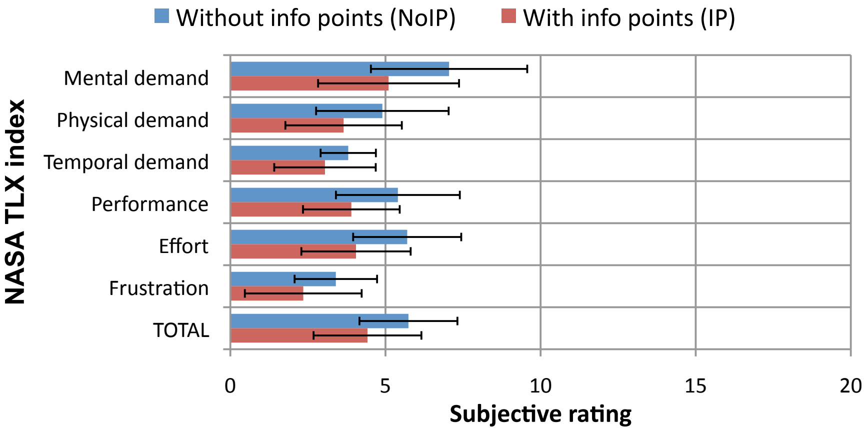


Figure 8.15: The average self-reported workload for each condition. We report all indices of the NASA TLX questionnaire, on a scale from 0 (low) to 20 (high), and the total workload: TOTAL is the weighted sum as defined in the NASA TLX instructions. The bars represent 95% confidence intervals.

Some participants also used more complex reasoning on the instructions and the structure of the building to recover from an error. For example, one participant who made a navigation error commented in the later interview: "In this case you reason more, like if you have to do 50 steps, you pick the longest corridor." Similarly, another participant got confused and rolled back to a walking instruction with a large number of steps to perform, returning to the beginning of a long corridor that she remembered. We observed that the turns given in the list of instructions also help in solving navigation ambiguities: in a few cases, we observed that when the overlay arrow pointed in the wrong direction users relied on the turn instruction visible in the instruction view.

Overall, the presence of info points improved performance on the navigation tasks, keeping users more on track and reducing the number of navigation errors. In the case of navigation errors, users often tried to match the list of instructions with prominent nearby landmarks, and to recover from these points.

Workload. A Pearson correlation test shows that the TLX results (the weighted sum of all TLX indices) have a positive correlation to the number of navigation errors. This correlation is statistically significant ($r = .563$, $N = 40$, $p < .01$). As shown in Figure 8.15, there is a slight tendency for a lower self-reported workload in the IP condition as compared to the NoIP condition. This difference is not statistically significant, therefore our hypothesis (H3) is not statistically supported. Furthermore, as the workload correlates positively with the number of errors, this difference in workload can be explained by the fact that participants made more errors in the NoIP condition.

All participants were asked if they felt a difference in completing the task with or without the info points. Three participants answered that it was the same with or without them. Yet, other participants claimed that it was “easier”, “more intuitive”, “useful” and “reassuring” to have the info points. Overall, most participants valued info points, but info points had no impact on their perceived workload.

Info points. During the interviews we additionally questioned all participants on their perceived value of the info points. The answers from the participants reveal that the value of info points was twofold, both as overview and confirmation.

- *Overview.* The info points acted as spots where users could get an overview of the sequence of upcoming instructions. For example, two participants stated that they were matching landmarks (e.g., toilets, corridors) between the info point and the environment, and one of them added that the info point showed “more than one task, not only the next but also the one after the next”. Another participant said that on the info point “you can look at the upcoming path: you get an overview, not only an arrow”. One further participant stated that “you see the direction where you have to go, not only the arrow. You get a good image of where you are.” Overall, “you could do [the task] without the info points, but with the info points it was much more intuitive. You could see the way you had to go.” Participants also contrasted this amount of overview to the arrow visualisation, which only shows information about the current instruction.

- *Confirmation.* Confirmation information is the second most needed type of information in navigation tasks, after wayfinding information [106]. From the interviews, it emerges that info points acted as checkpoints to obtain confirmation that the participant was on track. Info points were “good to check the position where you are located”, said one participant. At the info points “you have a point where you’re sure you get information [about] where you are”, added another participant. As one further participant remarked, “I was more sure that I’m on the right place, because with the [info points] I get feedback.”

The median number of info points used for each task is 3, thus almost every participant used each info point to complete the task (some participants used the same info point more than once). The average duration of an info point session (from the moment in which the info point was detected to the moment in which the participant stopped pointing the camera at it) is 3.72 seconds (SD = 3.26). As the standard deviation is high, in Figure

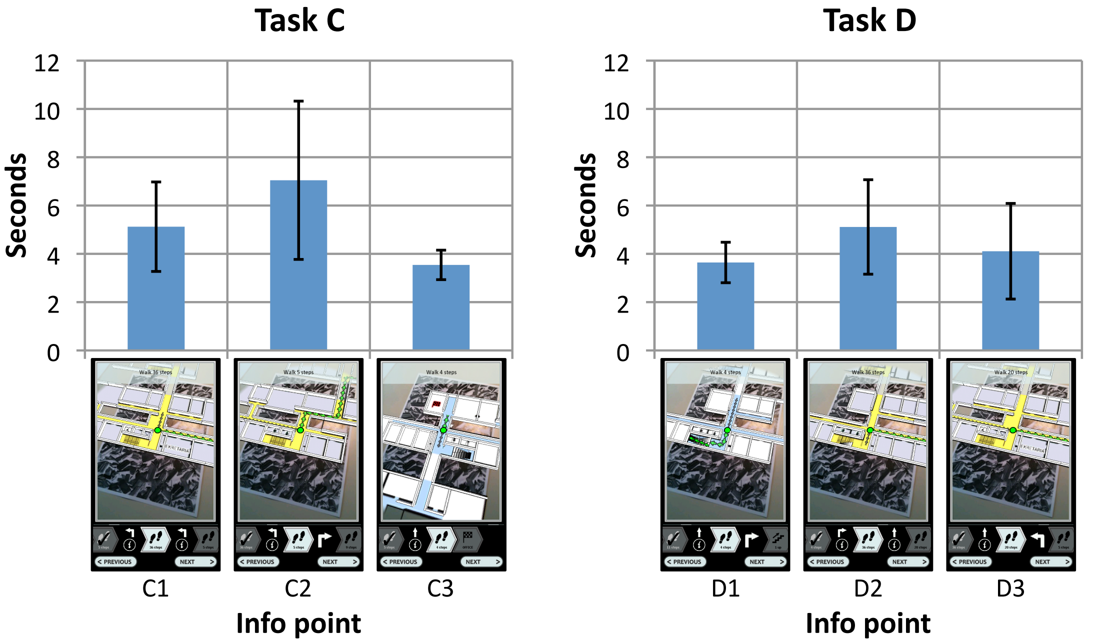


Figure 8.16: The average duration of info point sessions. We show timings and 95% confidence intervals for both tasks in the IP condition. Info points are presented according to the order in which they appear in the tasks (see Figure 8.12).

8.16 we look at the average duration of a session for each separate info point. Differences are not statistically significant, but they allow for interesting qualitative considerations. Usage sessions of info points where a turn was needed (C1, C2, D2) were longer than the info points where no turn was required (C3, D1, D3). In particular C2 took the longest. We reason that this is because, in contrast to C1 and D2, C2 was not followed by a single long sequence of steps but by a set of smaller activities, and therefore the info point presented a larger amount of information. In general, all sessions were only a few seconds long, supporting the observation that info points were used mainly as confirmation spots. This is in particular true for info points where users were not required to turn, whereas the info points at more complex intersections might present a higher component of overview.

8.2.3 Discussion

Overall, our interface was validated as an effective means to support indoor navigation with AR. As hypothesised from our previous experience with Signpost, the presence of info points causes an improvement in the performance of navigation tasks, significantly reducing the step deviation from the optimal path and contributing to a reduction in the number of navigation errors. While participants did not perceive a reduction in workload when info points were present, they valued them for their twofold role of providing confirmation that they were on track and showing an informative overview on the next instructions.

The role of confirmation was already present in Signpost, whereas overview is a novel role supported by the augmentations available at info points. We observed that participants often recovered from the navigation errors by looking for matches between the visualisation and the structure and landmarks in the surrounding building. The overview provided by info points was helpful in supporting this matching, as also stated by participants in the interviews.

Participants also contrasted the overview at info points with the poor overview of the arrow visualisation, which only shows information about the upcoming instruction. This informs our next design iteration: continuously supporting the matching between the instructions visualised on the screen and the environment is fundamental to help users monitoring their position in the building. This hints towards a more informative visualisation in between the info points, with similar characteristics to the visualisation at the info points. As proposed also by Butz et al. [19], increasing the information visualised when the localisation accuracy decreases makes the system more robust to navigational errors, because the user's spatial reasoning compensates for the lower accuracy of the system.

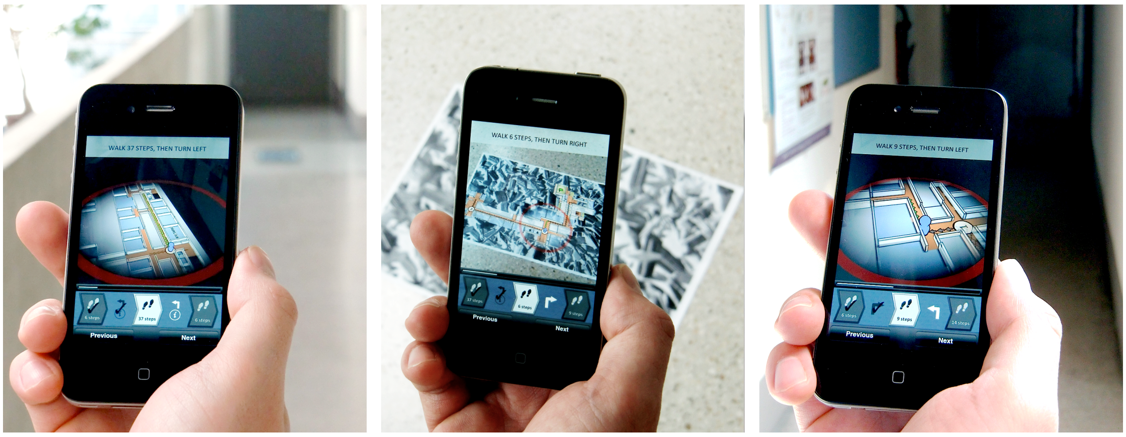
8.3 Sparse localisation and Mixed-Reality Cues

In light of the results of the previous experiment, we redesigned our interface with the objective of providing WIM-based navigation support continuously, rather than only at info points. To achieve this, we propose a smooth transition inside the MR space, depending on whether localisation is available or not, between AR (at info points) and VR (in-between info points). Our new design is shown in Figure 8.17.

8.3.1 Interface design

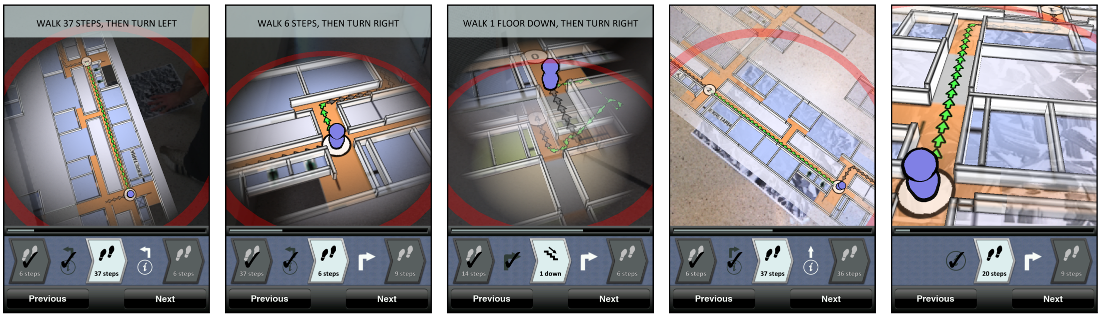
As in the previous design, we divide our interface into two main areas. The *instruction view* presents users with a list of instructions that must be performed in order to reach the next info point, or the target destination.

We redesigned the AR view as an *MR view* (Figure 8.17, upper part of the screen), which supports users' navigation with MR WIM views. The WIM contains some of the


WALK 37 STEPS, THEN TURN LEFT
WALK 6 STEPS, THEN TURN RIGHT
WALK 9 STEPS, THEN TURN LEFT
Previous
Next

(a) (b) (c)

Figure 8.17: Screenshots of our MR WIM views. We use MR to continuously support user navigation: as a user walks, we illustrate the next navigational instruction using VR (a). Once the user reaches an info point we show the whole path aligned with the environment using AR (b). After departing from the info point, we resort again to VR (c).


WALK 37 STEPS, THEN TURN LEFT
WALK 6 STEPS, THEN TURN RIGHT
WALK 1 FLOOR DOWN, THEN TURN RIGHT
6 steps
37 steps
6 steps
9 steps
14 steps
1 down
36 steps
20 steps
Previous
Next

(a) (b) (c) (d) (e)

Figure 8.18: WIM views in our interface. To provide continuous support for navigation with sparse localisation, we transition within the MR space: we use VR when there is no localisation (a-c) and AR when there is localisation (d-e).

landmarks of the building: walls, bridges, stairs, toilets and offices. Info points are also visualised. The WIM thus acts as a visual aid for the path that must be navigated, so that users can match their view on the physical environment with the view on the screen. We adapt the visualisation of the WIM depending on the current localisation accuracy.

If a user is walking between two info points, we present only information that details the current instruction in VR (Figure 8.18 (a-c)). In this case, our WIM visualisation is centred on the current instruction, so that we prominently present the next path segment to be traversed and the upcoming turn. If the user switches to the next instruction, a short animation moves the user's avatar (a small blue pawn) through the path segment to the next decision point, and the view is then centred on the next instruction. The same applies if the user switches to the previous instruction. The viewpoint position is fixed – as users are walking we do not know their location, therefore we assume that they are on the current instruction's path segment and we maintain the focus only on such segment. Yet, users control the angle at which the WIM is viewed by tilting the phone, so that parallax effects can help better understanding the 3D structure of the path (e.g., Figure 8.18 (a) shows a top-down view, while Figure 8.18 (b-c) show a more tilted view). Text information complements the VR view of the WIM and the rendering of the path, providing verbal details on the current instruction. Live video rendering is darkened – rather than disabled – to allow for camera targeting when a user approaches an info point.

When users approach an info point and they target it with the phone's camera, we provide more detailed information (Figure 8.18 (d-e)). In this case, we transition from the VR view to an AR view that provides an overview of the whole path – similar to the previous design – highlighting the current position, the path to the destination, and the already traversed path.

8.3.2 Controlled study

We have already validated the combination of AR with turn-by-turn instructions in our previous experiment. In this experiment, we focus on validating our refined interface design and evaluating if the WIM effectively supports users in monitoring their position continuously, not only at info points.

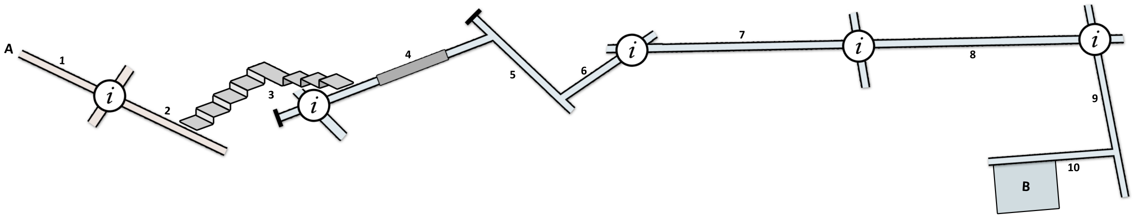


Figure 8.19: The path designed for the experiment. Participants walked the path from A to the office B. We divided the path into 10 path segments (1–10), which contain one floor change (3), one bridge (skyway) crossing (4) and a number of intersections.

Task. Participants were asked to navigate a significantly complex path within our department's building (Figure 8.19) – a task requiring one floor change, crossing a skyway (bridge-like connection between buildings), 7 changes in heading and 9 wayfinding decisions at intersections.

Research question. We wanted to see if participants continuously use the WIM views to monitor their position in the environment.

Experimental design. We adopted a *think-aloud* approach: we asked participants to state aloud all their navigational decisions and why they took them, during navigation. By choosing think-aloud, we aimed at collecting information on how users exploit environmental and interface cues during the task to make navigational decisions. We also tried to avoid the influence of post-task reasoning, which would have emerged more prominently with a post-task interview. At the end of the task, we collected usability data using the System Usability Scale (SUS) questionnaire [17].

Participants. 8 participants took part in the experiment, 4 male and 4 female, aged between 24 and 30. Each of them was compensated with a voucher for a local bookstore.

Apparatus. Our interface runs on an iPhone 4 at interactive frame rates. In VR mode, gyroscope, magnetometer and accelerometer data are fused with a linear Kalman filter to estimate the orientation of the device. Our implementation uses GLSL ES shaders for rendering.

Results. All participants successfully completed the task. The average score from SUS (on the range 0–100, 100 being the highest usability) is 75.31 (median = 83.75, SD = 20.29). The high standard deviation is inevitable, due to the small sample size and the subjective nature of the questionnaire. The result shows that participants did not have major usability issues with our system.

We transcribed all justifications of navigational decisions made by participants and divided them into 10 groups, one for each path segment. We then extracted all the keywords related to navigation (e.g., “turn left”), landmarks (e.g., “door”) or other spatial reasoning (e.g., “dead end”). Finally, for each path segment, we counted the utterances of each keyword. In Table 8.2, we distinguish between keywords that appear in the interface (top) and those that do not appear in the interface (bottom). The results of this process are also presented as tag clouds in Figure 8.20. Clearly, keywords cluster in proximity of corresponding building elements.

As expected, a large part (81%) of the keywords used by participants appears in textual form in the turn-by-turn instructions. In our previous experiment, we observed that turns and step counts are strong navigation indicators: our results from this experiment confirm this finding. Info points also act as strong landmarks: the posters (and the relative ID)

	Path segment										Total	
	1	2	3	4	5	6	7	8	9	10		
steps	4	2		6	3	2	5	1	3	3	29	**94**
straight	3	1					5	3	1		13	
turn right				5		3		4	5		17	
turn left		3	5		2						10	
up		2	7								9	
office										3	3	
info point	2		1			4	3	3			13	
intersection	1					1					2	**22**
stairs		3	2								5	
door		2								1	3	
bridge				2							2	
dead end				2	2						4	
corridor							1				1	
long distance							3				3	
map		1			1						2	

Table 8.2: Keyword utterances used by participants to justify navigational decisions. We count how many participants uttered a keyword on each path segment and in total. We further divide the keywords between those that appear written in the interface (top) and those that do not (bottom).

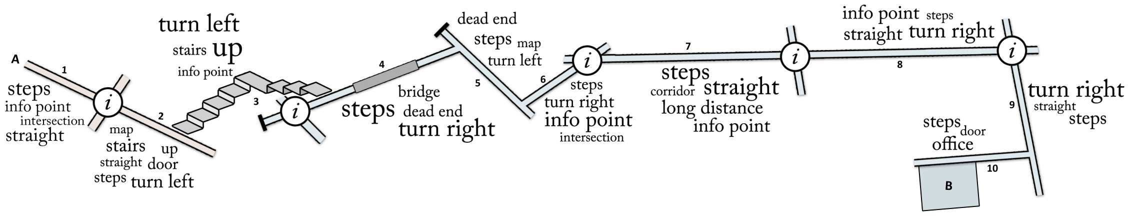


Figure 8.20: Keyword utterances as a tag cloud. We represent the utterances from Table 8.2 as tag clouds, for each path segment.

appear prominently on the floor of the building, in the WIM and in the instructions view. Some participants referred to them stating the specific number (e.g., “I am at info point 2”). The exact term “info point” did not appear in the interface, but the examiner used it in the introductory phase of the experiment. Nevertheless, one participant called it aptly a “check point”, hinting at its role for confirmation and overview.

The remaining part (19%) of the keywords relates to elements of the building that also appear in the WIM, but not textually in the interface: an *intersection* of two corridors, the *stairs*, the *door* of the staircase and of the target office, a *bridge* between two buildings, and a *dead end* in the corridor. A more complex keyword was used to match a *long distance* to travel (37 steps) with the one sufficiently long *corridor* in the apparent surroundings. Together with the keyword “dead end”, this points at the visual support of the WIM for excluding impossible routes. Interestingly, the *map* that we used as floor texture for the

WIM was also used twice.

8.3.3 Discussion

While keywords from turn-by-turn instructions are prominent in participants' justifications, the results suggest that there was also an underlying process, supported by the WIM, of matching the environment with the interface. Further, keyword utterances occurred throughout the path, not only at info points. Many auxiliary landmarks (e.g., plants, fire extinguishers, tables) were missing in our WIM visualisation: indeed, such landmarks also did not appear as keywords in participants' statements.

Our evaluations confirms the validity of our design – combining MR WIM views with turn-by-turn instructions can support indoor navigation when continuous localisation is not available. Furthermore, our results suggest that providing informative views – in the form of a WIM – supports users in monitoring their position by matching the view on the interface with the physical elements of the environment. In line with our previous observations, our results show that users' navigational ability can be exploited to substitute continuous localisation, if a sufficiently informative and consistent interface is provided. An important aspect hereby is that transitioning within MR allows us to keep the WIM always on the screen and adapt affordances depending on the state of localisation.

8.4 Reflection on our research questions

The results presented in this chapter show that AR can be successfully used to support indoor navigation at critical decision points in the building, with minimal instrumentation of the environment by means of simple info points.

If we consider the results of this chapter from the perspective of our main research questions (Section 1.3), we see that AR succeeds in supporting users at info point, with informative WIM views that can be intuitively browsed by moving the phone over the info point. We find an analogy with the results presented in Chapter 6, where users also quickly grasped the interaction metaphor of browsing augmented maps.

Similarly to the results from Chapter 7, we see that an arrow-based AR visualisation is not sufficiently informative. In this case, AR fails to support navigation, because the egocentric information does not provide sufficient overview and is not always reliable, due to the limitations of tracking technology. Our results show that AR can be complemented by a VR visualisation that is more robust to tracking inaccuracies, and provides more details needed by users during the wayfinding task.

In summary, the results from our experiments suggest two key design recommendations. First, using augmented maps at difficult decision points in the building is a good design choice for improving navigation performance. Since maps are typically available in many buildings, this design can be easily integrated with the already existing infrastructure. Second, using detailed egocentric AR cues is not a good design choice when the navigation system has a high uncertainty on the user's position and orientation. Rather, using more informative views (such as our VR WIM) is a good design choice that helps people compensating for the system's uncertainty. In relation to the second recommendation, interface designers should consider that we only evaluated two extreme cases: in one case we had perfect tracking and localisation (at the info point), in the other case we had no localisation at all (in between info points). Some navigation systems use intermediate localisation technology – for example, dead reckoning, which offers continuous localisation but does not have a good accuracy. For these intermediate cases, our results and the related work on the topic [19, 50] are consistent in suggesting that a good design choice is to adapt at runtime the amount of information visualised, in function of the current localisation accuracy.

PART V

Conclusion

Chapter 9

Discussion and future directions

This goal of this thesis was to investigate the usability of AR in the context of handheld navigation systems. In particular, our focus was on finding *when AR enhances handheld navigation systems* and, *when it fails to enhance them, how we can complement AR with other interfaces* to build a hybrid and more effective interface.

In order to design clear and informative experimental evaluations, we subdivided the problem into sub-tasks and studied each of them separately. We organised our research work in two separate parts, one focused on *exploration* and another one on *wayfinding*, observing the duality of the two tasks in related literature on human navigation. We further subdivided exploration into egocentric (Chapter 5) and exocentric (Chapter 6), depending on how AR is used to present information, and separated outdoor wayfinding (Chapter 7) and indoor wayfinding (Chapter 8). We studied the four tasks separately and presented our results in the corresponding chapters.

In this final part of the thesis, we give an overview of all our results from a global point of view, grouping together all the lessons learned and forming a concise set of general considerations. These can be useful as high-level guidelines for interface designers who want to integrate AR into handheld navigation systems. Finally, we conclude the chapter highlighting the most promising directions for future research on the topic.

9.1 Lessons learned

The related work suggests that a key role of AR in enhancing navigation systems is the intuitive and natural interaction metaphor that it provides. First of all, our experimental results confirm and strengthen this consideration.

- AR provides an intuitive means for *pointing at information in the environment.* Both in Chapter 5 and in Chapter 7, we observed that users quickly grasp the concept of pointing the device to different directions in the environment, in order to access the information.

- AR provides an intuitive means for *browsing paper-based artefacts.* Both in Chapter 6 and in Chapter 8, we observed that users quickly grasp the concept of sweeping the device around a paper map or a poster, in order to change the point of view on the information.

These first two considerations inform us about the general usability of AR in handheld navigation systems, and they confirm that the interaction metaphor offered by AR is found to be intuitive by end users. We further identified two specific scenarios, in which AR is not only usable but also enhances handheld navigation systems.

- In exploration, AR is advantageous for supporting *discussion and common understanding on the information space* (in our case, a paper map). As we observed in Chapter 6, the availability of a physical prop encourages collaborative behaviour and discussion, in contrast to a handheld navigation system that does not use AR. It must be also noted how *the availability of one device for each user supports easier understanding of the information* with less communicational effort: this stresses the necessity to make the augmented information visible and accessible to all parties involved in the discussion.
- In wayfinding, AR support is needed *when decisions have to be made, in particular at intersections and in the proximity of a destination.* Both in Chapter 7 and in Chapter 8, we observed how AR cues can have a *dual role of providing navigational information as well as confirmation of being on track.* Reflecting on the first two points made at the beginning of this section, AR support is not limited to egocentric cues augmented on the environment (as in Chapter 7), but it also extends to exocentric overviews that can be browsed intuitively (as in Chapter 8).

In the course of this thesis, we also identified a number of scenarios in which AR potentially fails to enhance handheld navigation systems, unless it is complemented by other interfaces. From the design iterations and the evaluations we conducted, we can draw a few considerations on this matter.

- In egocentric AR, users need *guidance when augmentations are outside the camera view.* We did not find a unique interface that can be used to always provide this guidance, but we rather observed that the choice of such interface depends on the task users are engaged in. In Chapter 5, we observed that *overlays can be used for pointing users to one single off-screen augmentation*, exploiting the intuitiveness of pointing in AR. A similar overlay was also successfully used in Chapter 7. In Chapter 5, we also found that *transitioning to other interfaces is more efficient than using an overlay, when overview on all augmentations is needed.* In our case, we found that both an egocentric view (a 360° panorama) and an exocentric view (a top-down map) can provide this overview.

- In exocentric AR, we found that bimanual *coordination of the device with the paper map is difficult while walking.* In Chapter 6, we observed that usage of the handheld navigation system while walking is still possible, provided that *the system transitions to non-AR interfaces if AR is not necessary for the user's task.* In our case, we use AR for exploring and selecting photographs from a paper map, and a simple thumbnail list for looking at the photograph. We adopt this concept also in Section 8.3, where a number of public AR information points are distributed in a building: users can explore overview information in AR while standing at the information points, or take it away with them in a VR mode.

- Overall, interface designers also need to consider that the adoption of AR for a specific navigational task is constrained by the *support given by more established interfaces for the same task.* It is not only important to evaluate if AR enhances navigation tasks, but also to consider if more established interfaces cannot support such tasks. For example, in Chapter 7 we saw that for simple outdoor navigation tasks users are reluctant to abandon a map-based interface – an established interface that works sufficiently well – for a novel interface like AR.

Finally, two main themes touched all the work in this thesis. The first is *the importance of considering tracking at all stages of the design process* of a handheld navigation system. With this, we do not only refer to the understanding that good tracking is necessary for good AR, which is an established consideration within the AR community. We also refer to the need of not disregarding tracking inaccuracies in the interface design. For example, we observed both in Chapter 7 and in Chapter 8 that arrow-based augmentations create expectations of high precision in the registration, and when these expectations are broken, users are left disoriented. This clearly points to the need of actively communicating the inaccuracies in the interface design, in order to avoid misleading the users. Human abilities can be exploited for filling the gaps in which tracking is too inaccurate or impossible. For example, in Chapter 8 we show a system that transitions from AR to more informative views when tracking is not available. Finally, the role of tracking as a factor must also be considered in the user evaluations. For example, in Chapter 6 we observed that a revised tracking technology changed the users' behaviour with augmented maps, making the usage more spontaneous and agile. Overall, throughout our research work we found that tracking must be considered not only a key technological requirement, but also a fundamental factor in the user experience of the application. Tracking should therefore be considered not only during implementation but also in the design phase.

The second common theme is the need to *find the most effective evaluation techniques* for mobile AR navigation systems. In line with common HCI practices, our experience shows that a triangulation of different methodologies is key to understand how people benefit from mobile AR. We gained insight on users' performance and preferences from quantitative data in controlled studies, while qualitative data let us better understand why differences or similarities occurred in the quantitative data. Real-world deployment gave us further insight on how the results from the studies map to everyday life, when the environment, the users' tasks and the users' priorities are more varied and variable than in a controlled study. Overall, we found that a key factor in mobile AR evaluations is not only choosing the most effective evaluation techniques, but also how complementary techniques are combined to form a "big picture" of the usability of the AR system.

9.2 Future directions

We see three key future directions for the research work conducted in this thesis. The first and most straightforward direction is to closely follow new findings in tracking technology (see Section 2.4) and explore the support they can provide for novel interface designs. For example, the recent work by Arth et al. [3, 4] enables accurate 6 DOF localisation in urban scenarios on mobile phones and expands the possibilities for egocentric exploration in AR. The foreseeable future availability of a SLAM-like solution for mobile phones would allow full 6 DOF tracking, thus enabling more accurate AR cues also for wayfinding, where we noticed that users typically operate the interface while walking. Following the most recent trends in tracking and exploring how they can enable novel interface designs is a research direction in clear continuity with the work done in this thesis.

A correlated research direction is to expand our work with large-scale longitudinal studies. While this was not easily doable a few years ago, it is now becoming possible thanks to two key technological advancements: first, the availability of commercial tracking solutions, such as Vuforia [149], that allow easy development of AR applications with robust state-of-the-art tracking. Second, the relative ease in penetrating a very large user base by publishing these application on the app stores (see, for example, the recent work done by Henze et al. [54]). For example, it is feasible to implement an application that augments public maps with content retrieved from online sources, and to publish it on the app stores. Similarly to the study we presented in Section 8.1.3, anonymous usage logs can provide deeper insight on how users operate augmented reality in their daily life and over a long period of time. This type of insight is a strong quantitative complement to the results from the controlled studies presented in this thesis.

A final and more speculative research direction is to look at how AR can support novel ways of communicating navigational instructions: this will require combining AR interface design with a deep cognitive understanding of how humans navigate. In particular, the recent trends in handheld navigation systems (see Section 2.2) and our most recent work from Section 8.3 inspire us in looking at how landmark-centred instructions can enhance human navigation abilities, rather than replace them. Landmark-centred instructions are the way people typically communicate navigation instructions to each other, and AR could provide instructions in a similar fashion. For example, AR could be used to highlight an important landmark at a distance, so that the navigation system can use it for explaining to the user where she is supposed to turn. Overall, we envision that a key role of AR could be providing instructions in a natural way, which differentiates itself from the very abstract turn-by-turn instructions typical in handheld navigation systems. Similar to the partially explored role of photographs as navigation cues, the tight bound of AR to the real world can be exploited for providing natural instructions that are clearly anchored to landmarks in the environment. These instructions would be much closer to the way humans are used to navigate with their own minds, when no navigation system is at hand.

PART VI

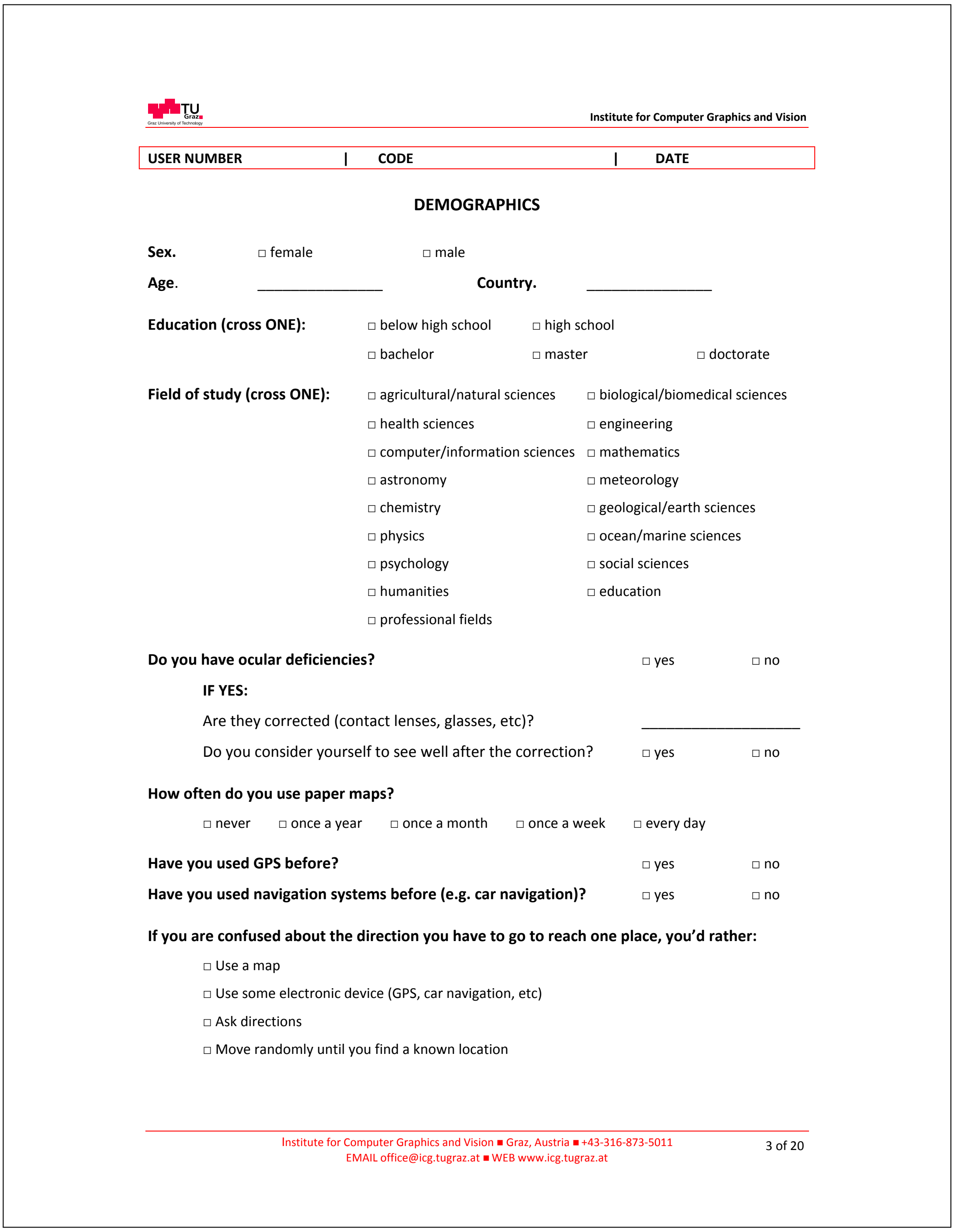
Appendix

Appendix A

Questionnaires

A.1 Pointing to a specific augmentation

This section presents the questionnaire used in the user study presented in Section 5.1.



USER NUMBER	CODE	DATE

DEMOGRAPHICS

Sex. □ female □ male

Age. ________________ **Country.** ________________

Education (cross ONE): □ below high school □ high school □ bachelor □ master □ doctorate

Field of study (cross ONE):

□ agricultural/natural sciences
□ health sciences
□ computer/information sciences
□ astronomy
□ chemistry
□ physics
□ psychology
□ humanities
□ professional fields
□ biological/biomedical sciences
□ engineering
□ mathematics
□ meteorology
□ geological/earth sciences
□ ocean/marine sciences
□ social sciences
□ education

Do you have ocular deficiencies? □ yes □ no

IF YES:

Are they corrected (contact lenses, glasses, etc)? ____________________

Do you consider yourself to see well after the correction? □ yes □ no

How often do you use paper maps?

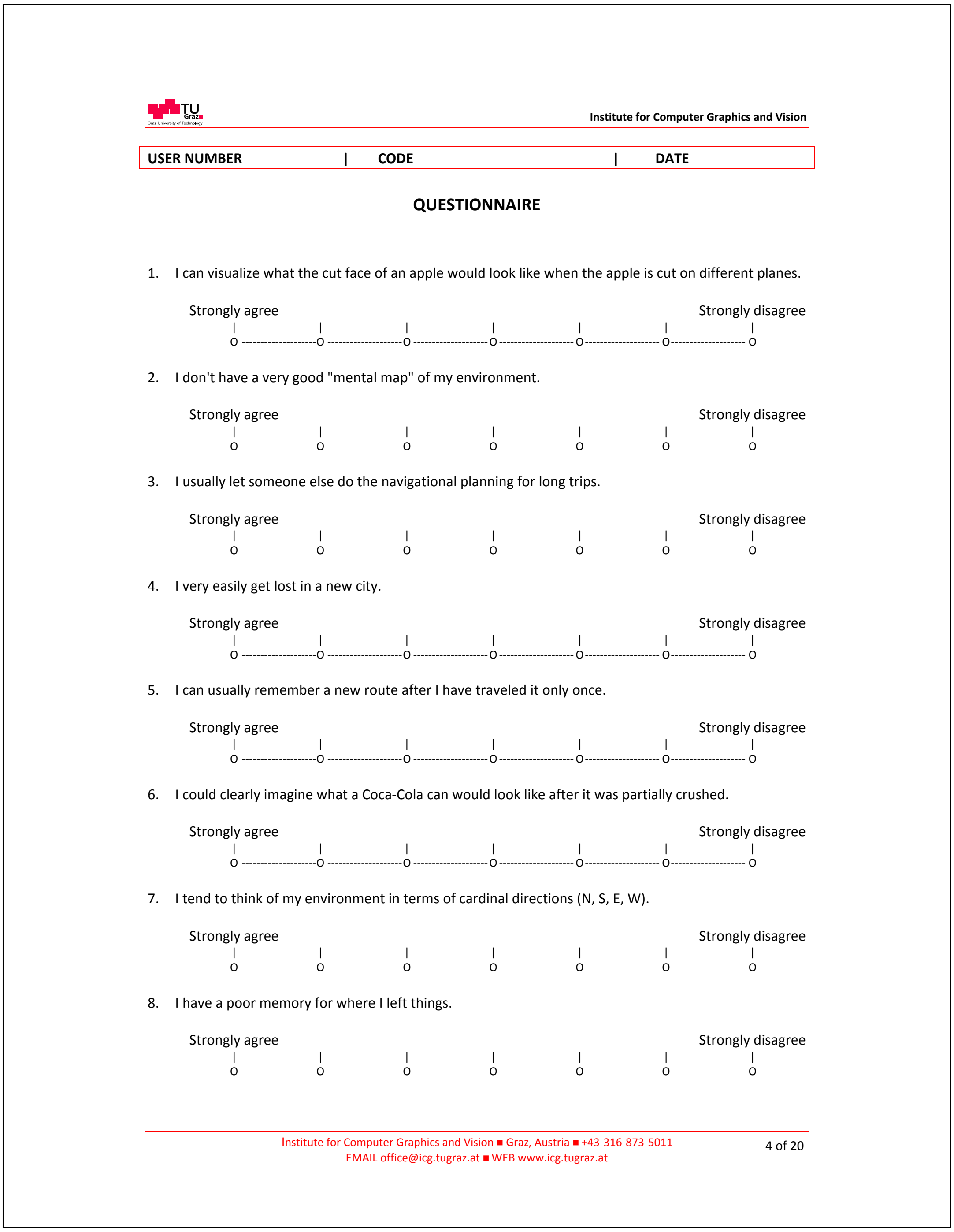
□ never □ once a year □ once a month □ once a week □ every day

Have you used GPS before? □ yes □ no

Have you used navigation systems before (e.g. car navigation)? □ yes □ no

If you are confused about the direction you have to go to reach one place, you'd rather:

□ Use a map
□ Use some electronic device (GPS, car navigation, etc)
□ Ask directions
□ Move randomly until you find a known location



TU Graz Graz University of Technology

Institute for Computer Graphics and Vision

USER NUMBER | CODE | DATE

QUESTIONNAIRE

1. I can visualize what the cut face of an apple would look like when the apple is cut on different planes.

Strongly agree Strongly disagree

| | | | | | |

O --------------------O --------------------O --------------------O --------------------O--------------------O--------------------O

2. I don't have a very good "mental map" of my environment.

Strongly agree Strongly disagree

| | | | | | |

O --------------------O --------------------O --------------------O --------------------O--------------------O--------------------O

3. I usually let someone else do the navigational planning for long trips.

Strongly agree Strongly disagree

| | | | | | |

O --------------------O --------------------O --------------------O --------------------O--------------------O--------------------O

4. I very easily get lost in a new city.

Strongly agree Strongly disagree

| | | | | | |

O --------------------O --------------------O --------------------O --------------------O--------------------O--------------------O

5. I can usually remember a new route after I have traveled it only once.

Strongly agree Strongly disagree

| | | | | | |

O --------------------O --------------------O --------------------O --------------------O--------------------O--------------------O

6. I could clearly imagine what a Coca-Cola can would look like after it was partially crushed.

Strongly agree Strongly disagree

| | | | | | |

O --------------------O --------------------O --------------------O --------------------O--------------------O--------------------O

7. I tend to think of my environment in terms of cardinal directions (N, S, E, W).

Strongly agree Strongly disagree

| | | | | | |

O --------------------O --------------------O --------------------O --------------------O--------------------O--------------------O

8. I have a poor memory for where I left things.

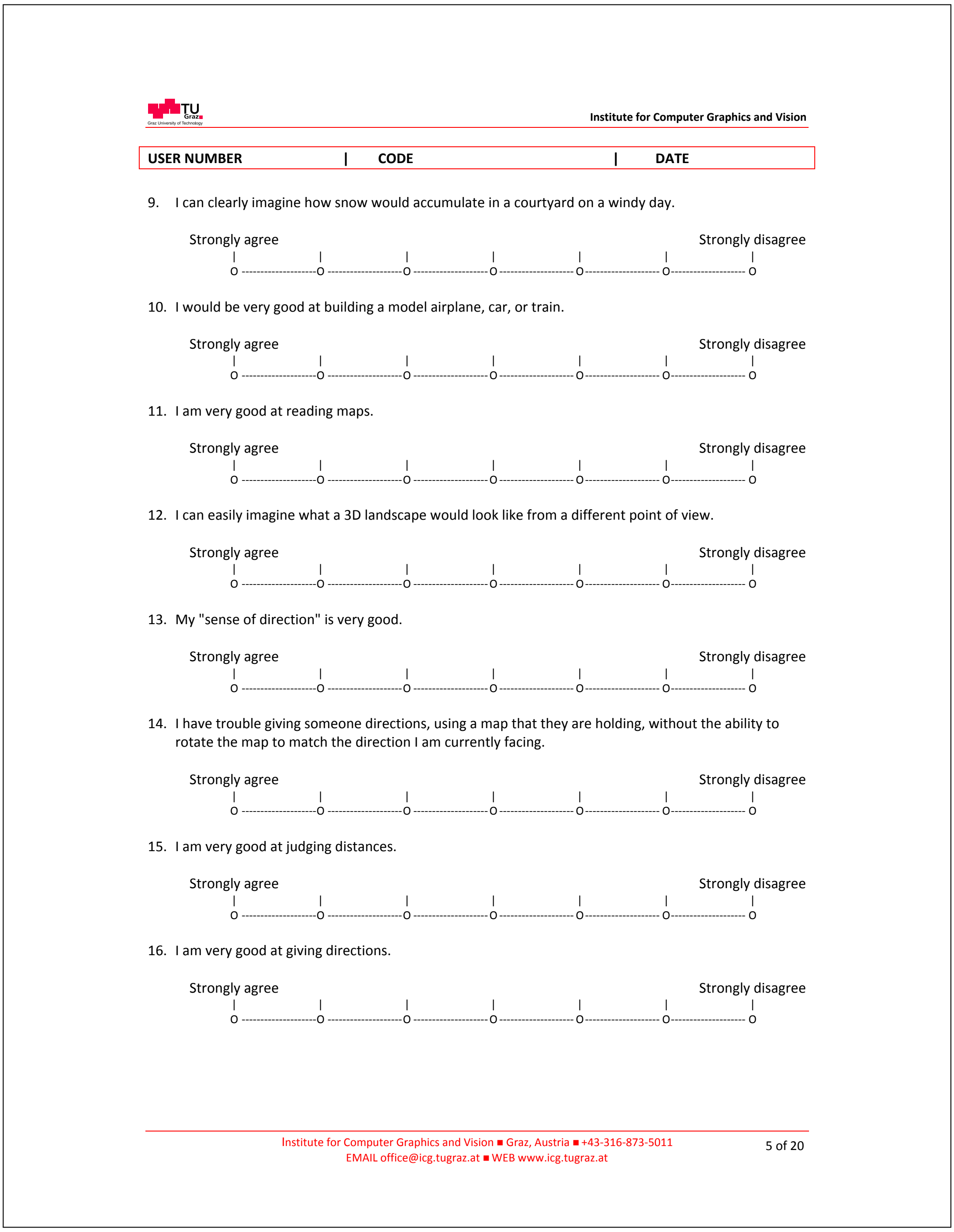
Strongly agree Strongly disagree

| | | | | | |

O --------------------O --------------------O --------------------O --------------------O--------------------O--------------------O

Institute for Computer Graphics and Vision ■ Graz, Austria ■ +43-316-873-5011
EMAIL office@icg.tugraz.at ■ WEB www.icg.tugraz.at

4 of 20



Institute for Computer Graphics and Vision

USER NUMBER	CODE	DATE

9. I can clearly imagine how snow would accumulate in a courtyard on a windy day.

Strongly agree Strongly disagree

O --------------------O --------------------O --------------------O --------------------O --------------------O --------------------O

10. I would be very good at building a model airplane, car, or train.

Strongly agree Strongly disagree

O --------------------O --------------------O --------------------O --------------------O --------------------O --------------------O

11. I am very good at reading maps.

Strongly agree Strongly disagree

O --------------------O --------------------O --------------------O --------------------O --------------------O --------------------O

12. I can easily imagine what a 3D landscape would look like from a different point of view.

Strongly agree Strongly disagree

O --------------------O --------------------O --------------------O --------------------O --------------------O --------------------O

13. My "sense of direction" is very good.

Strongly agree Strongly disagree

O --------------------O --------------------O --------------------O --------------------O --------------------O --------------------O

14. I have trouble giving someone directions, using a map that they are holding, without the ability to rotate the map to match the direction I am currently facing.

Strongly agree Strongly disagree

O --------------------O --------------------O --------------------O --------------------O --------------------O --------------------O

15. I am very good at judging distances.

Strongly agree Strongly disagree

O --------------------O --------------------O --------------------O --------------------O --------------------O --------------------O

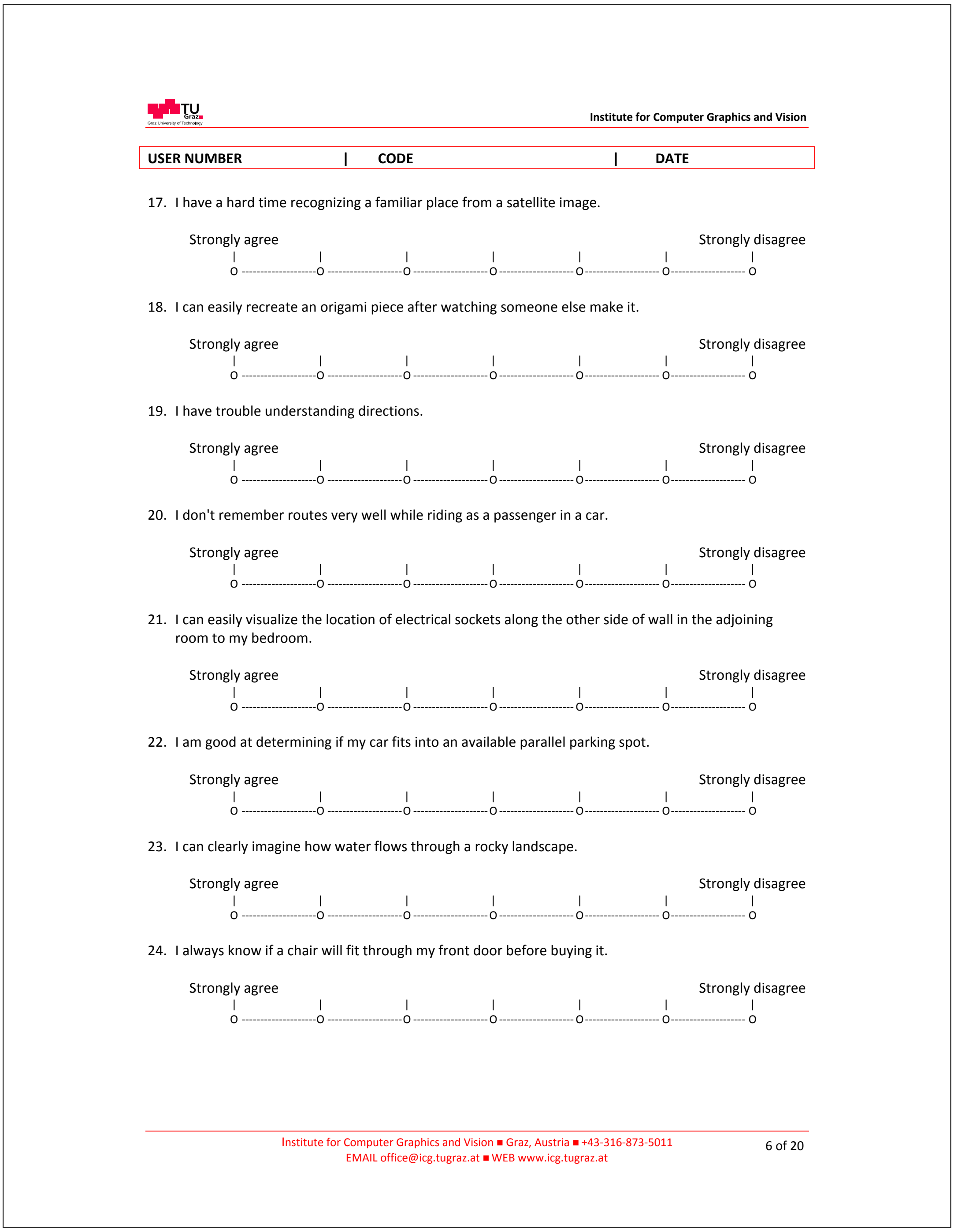
16. I am very good at giving directions.

Strongly agree Strongly disagree

O --------------------O --------------------O --------------------O --------------------O --------------------O --------------------O

Institute for Computer Graphics and Vision ■ Graz, Austria ■ +43-316-873-5011
EMAIL office@icg.tugraz.at ■ WEB www.icg.tugraz.at

5 of 20



TU Graz

Institute for Computer Graphics and Vision

USER NUMBER	**CODE**	**DATE**

17. I have a hard time recognizing a familiar place from a satellite image.

Strongly agree Strongly disagree

```
|              |              |              |              |              |              |
O -------------O -------------O -------------O -------------O -------------O ------------- O
```

18. I can easily recreate an origami piece after watching someone else make it.

Strongly agree Strongly disagree

```
|              |              |              |              |              |              |
O -------------O -------------O -------------O -------------O -------------O ------------- O
```

19. I have trouble understanding directions.

Strongly agree Strongly disagree

```
|              |              |              |              |              |              |
O -------------O -------------O -------------O -------------O -------------O ------------- O
```

20. I don't remember routes very well while riding as a passenger in a car.

Strongly agree Strongly disagree

```
|              |              |              |              |              |              |
O -------------O -------------O -------------O -------------O -------------O ------------- O
```

21. I can easily visualize the location of electrical sockets along the other side of wall in the adjoining room to my bedroom.

Strongly agree Strongly disagree

```
|              |              |              |              |              |              |
O -------------O -------------O -------------O -------------O -------------O ------------- O
```

22. I am good at determining if my car fits into an available parallel parking spot.

Strongly agree Strongly disagree

```
|              |              |              |              |              |              |
O -------------O -------------O -------------O -------------O -------------O ------------- O
```

23. I can clearly imagine how water flows through a rocky landscape.

Strongly agree Strongly disagree

```
|              |              |              |              |              |              |
O -------------O -------------O -------------O -------------O -------------O ------------- O
```

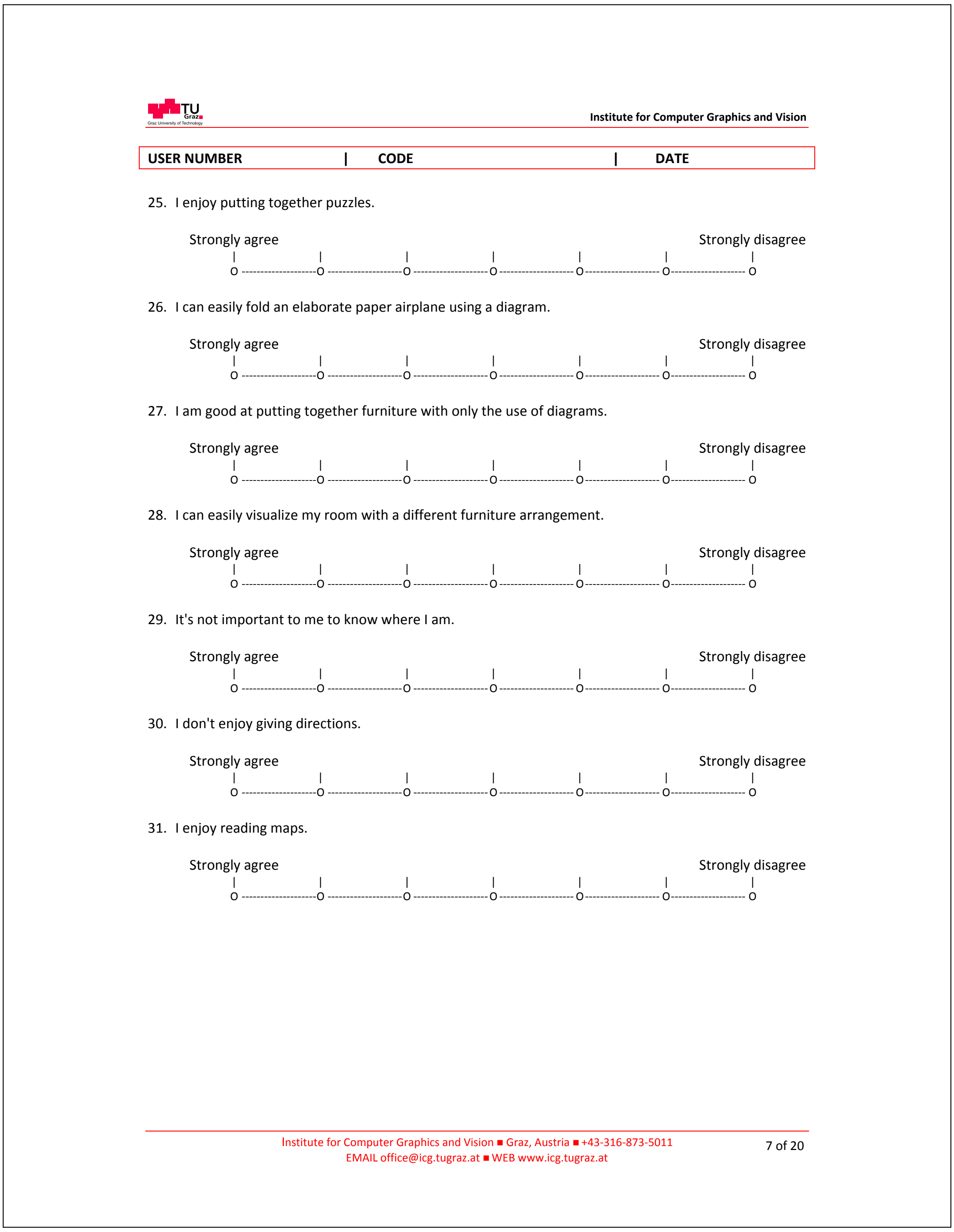
24. I always know if a chair will fit through my front door before buying it.

Strongly agree Strongly disagree

```
|              |              |              |              |              |              |
O -------------O -------------O -------------O -------------O -------------O ------------- O
```

Institute for Computer Graphics and Vision ■ Graz, Austria ■ +43-316-873-5011
EMAIL office@icg.tugraz.at ■ WEB www.icg.tugraz.at

6 of 20



TU Graz
Graz University of Technology

Institute for Computer Graphics and Vision

USER NUMBER | CODE | DATE

25. I enjoy putting together puzzles.

Strongly agree Strongly disagree
| | | | | | |
O -------------------O -------------------O -------------------O -------------------O -------------------O -------------------O

26. I can easily fold an elaborate paper airplane using a diagram.

Strongly agree Strongly disagree
| | | | | | |
O -------------------O -------------------O -------------------O -------------------O -------------------O -------------------O

27. I am good at putting together furniture with only the use of diagrams.

Strongly agree Strongly disagree
| | | | | | |
O -------------------O -------------------O -------------------O -------------------O -------------------O -------------------O

28. I can easily visualize my room with a different furniture arrangement.

Strongly agree Strongly disagree
| | | | | | |
O -------------------O -------------------O -------------------O -------------------O -------------------O -------------------O

29. It's not important to me to know where I am.

Strongly agree Strongly disagree
| | | | | | |
O -------------------O -------------------O -------------------O -------------------O -------------------O -------------------O

30. I don't enjoy giving directions.

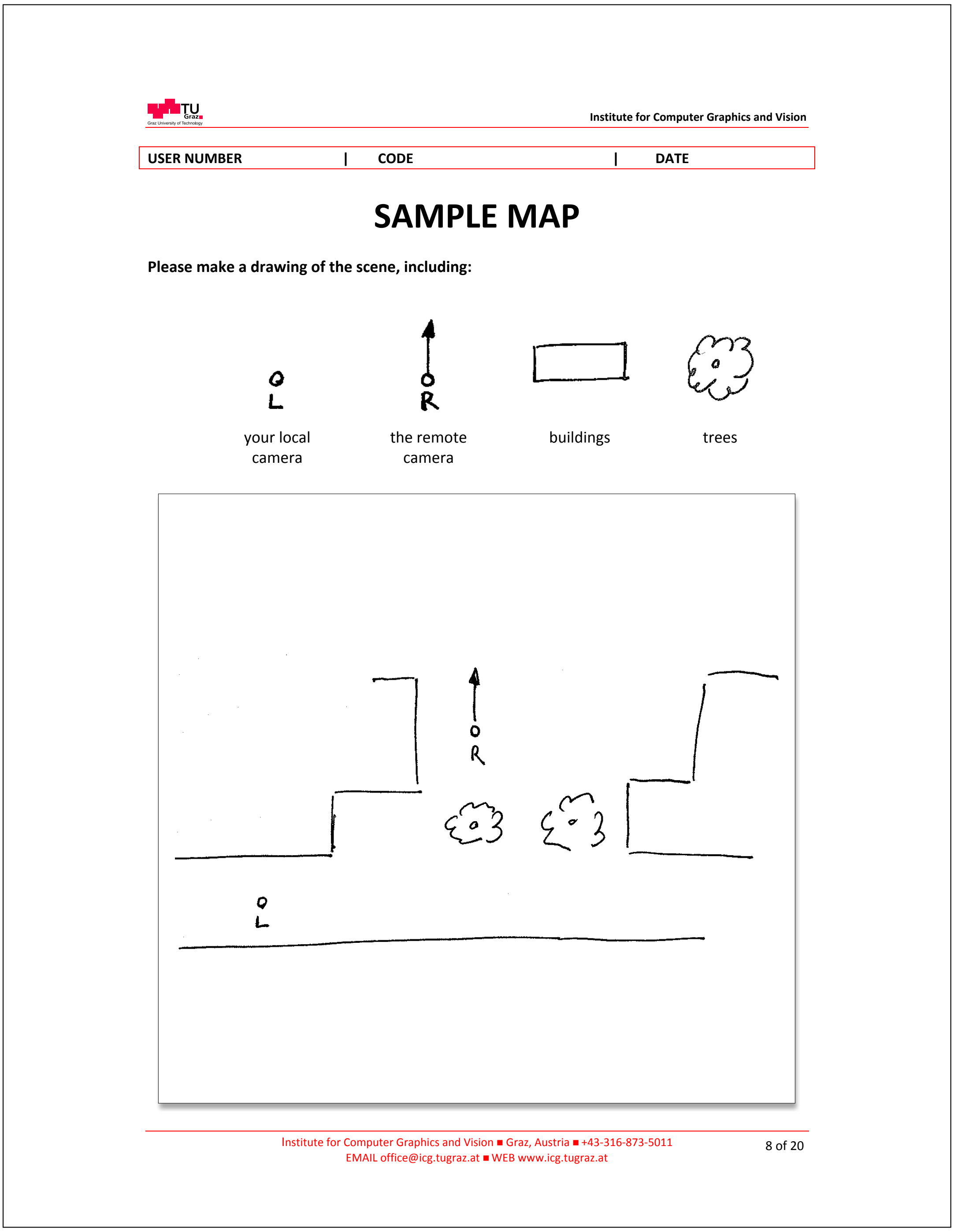
Strongly agree Strongly disagree
| | | | | | |
O -------------------O -------------------O -------------------O -------------------O -------------------O -------------------O

31. I enjoy reading maps.

Strongly agree Strongly disagree
| | | | | | |
O -------------------O -------------------O -------------------O -------------------O -------------------O -------------------O

Institute for Computer Graphics and Vision ■ Graz, Austria ■ +43-316-873-5011
EMAIL office@icg.tugraz.at ■ WEB www.icg.tugraz.at

7 of 20



Institute for Computer Graphics and Vision

USER NUMBER	CODE	DATE

SAMPLE MAP

Please make a drawing of the scene, including:

your local camera

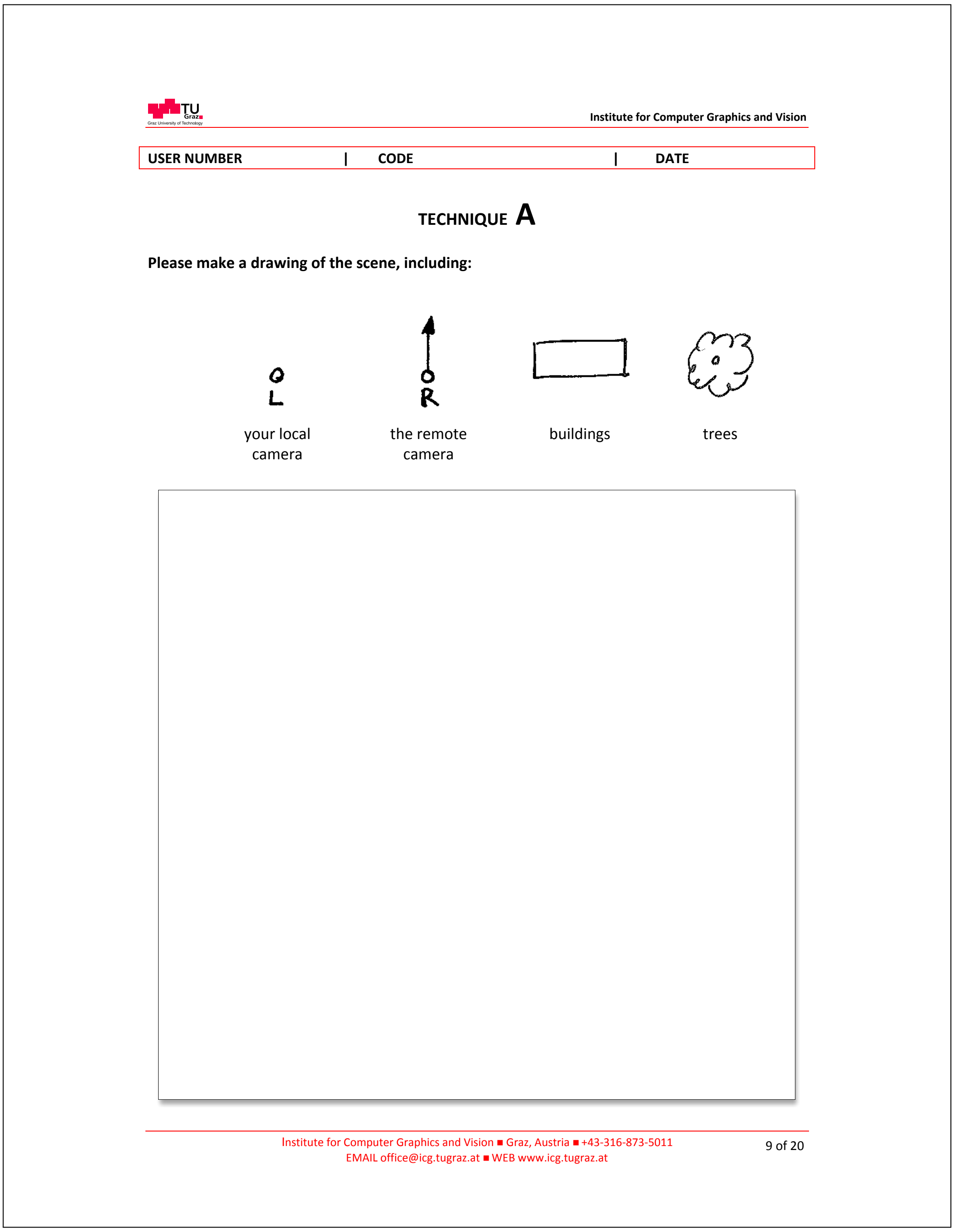
the remote camera

buildings

trees

Institute for Computer Graphics and Vision ■ Graz, Austria ■ +43-316-873-5011
EMAIL office@icg.tugraz.at ■ WEB www.icg.tugraz.at

8 of 20



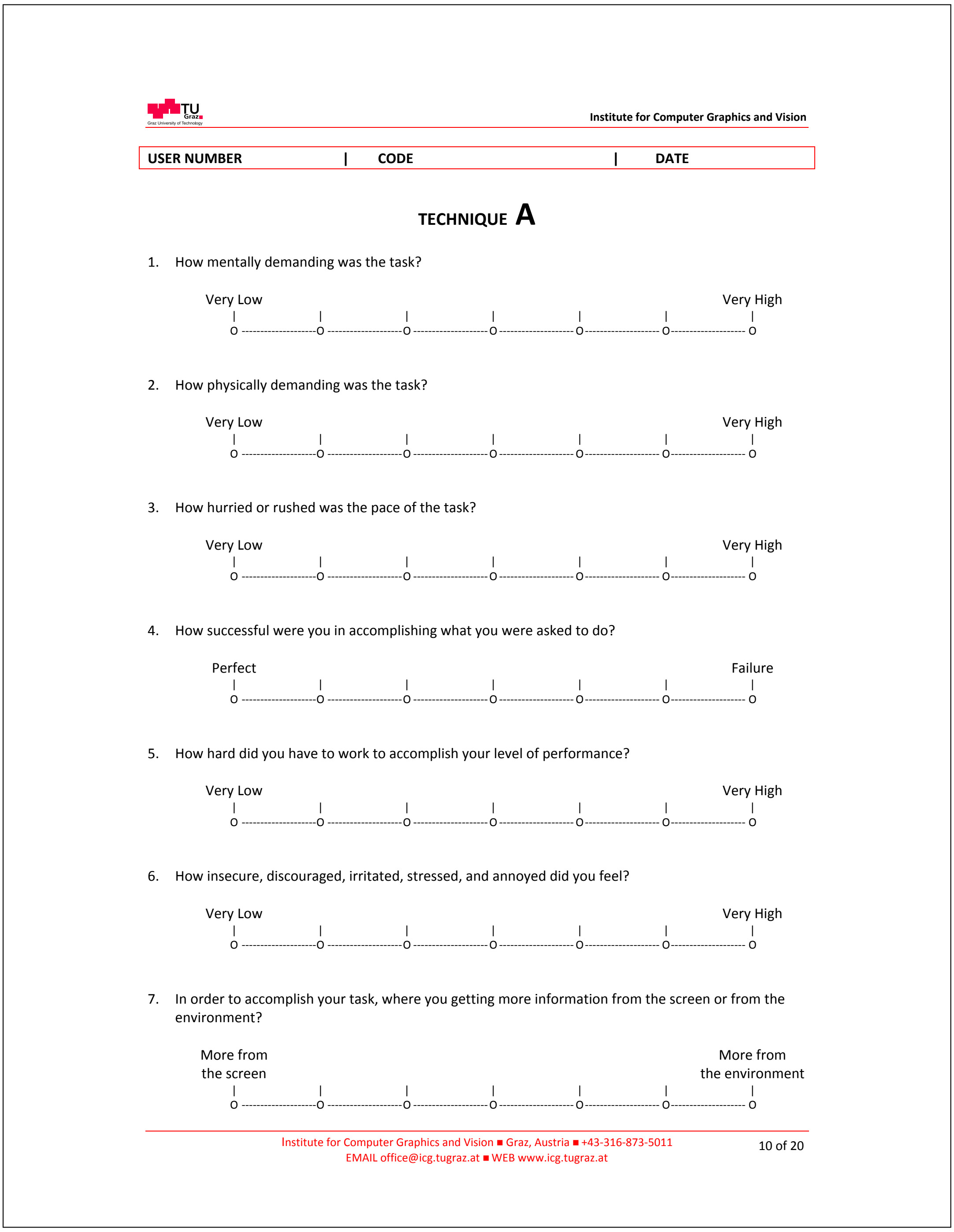
Institute for Computer Graphics and Vision

USER NUMBER | CODE | DATE

TECHNIQUE **A**

Please make a drawing of the scene, including:

your local camera	the remote camera	buildings	trees



TU Graz
Graz University of Technology

Institute for Computer Graphics and Vision

USER NUMBER | CODE | DATE

TECHNIQUE A

1. How mentally demanding was the task?

Very Low Very High

| | | | | | |

O --------------------O -------------------- O -------------------- O -------------------- O -------------------- O -------------------- O

2. How physically demanding was the task?

Very Low Very High

| | | | | | |

O --------------------O -------------------- O -------------------- O -------------------- O -------------------- O -------------------- O

3. How hurried or rushed was the pace of the task?

Very Low Very High

| | | | | | |

O --------------------O -------------------- O -------------------- O -------------------- O -------------------- O -------------------- O

4. How successful were you in accomplishing what you were asked to do?

Perfect Failure

| | | | | | |

O --------------------O -------------------- O -------------------- O -------------------- O -------------------- O -------------------- O

5. How hard did you have to work to accomplish your level of performance?

Very Low Very High

| | | | | | |

O --------------------O -------------------- O -------------------- O -------------------- O -------------------- O -------------------- O

6. How insecure, discouraged, irritated, stressed, and annoyed did you feel?

Very Low Very High

| | | | | | |

O --------------------O -------------------- O -------------------- O -------------------- O -------------------- O -------------------- O

7. In order to accomplish your task, where you getting more information from the screen or from the environment?

More from
the screen

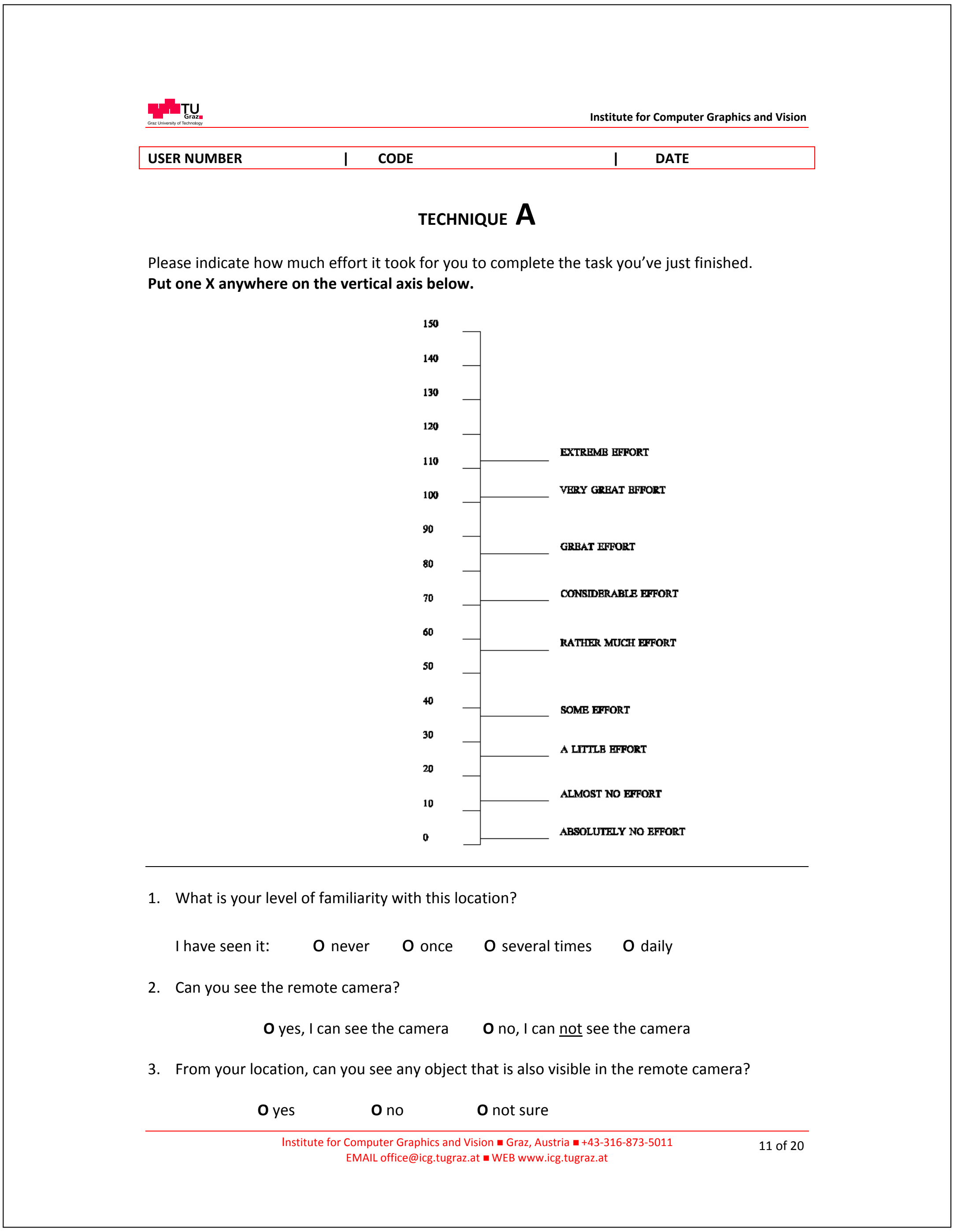
More from
the environment

| | | | | | |

O --------------------O -------------------- O -------------------- O -------------------- O -------------------- O -------------------- O

Institute for Computer Graphics and Vision ■ Graz, Austria ■ +43-316-873-5011
EMAIL office@icg.tugraz.at ■ WEB www.icg.tugraz.at

10 of 20



USER NUMBER | CODE | DATE

TECHNIQUE A

Please indicate how much effort it took for you to complete the task you've just finished.
Put one X anywhere on the vertical axis below.

150
140
130
120
110 EXTREME EFFORT
100 VERY GREAT EFFORT
90
80 GREAT EFFORT
70 CONSIDERABLE EFFORT
60 RATHER MUCH EFFORT
50
40 SOME EFFORT
30 A LITTLE EFFORT
20 ALMOST NO EFFORT
10
0 ABSOLUTELY NO EFFORT

1. What is your level of familiarity with this location?

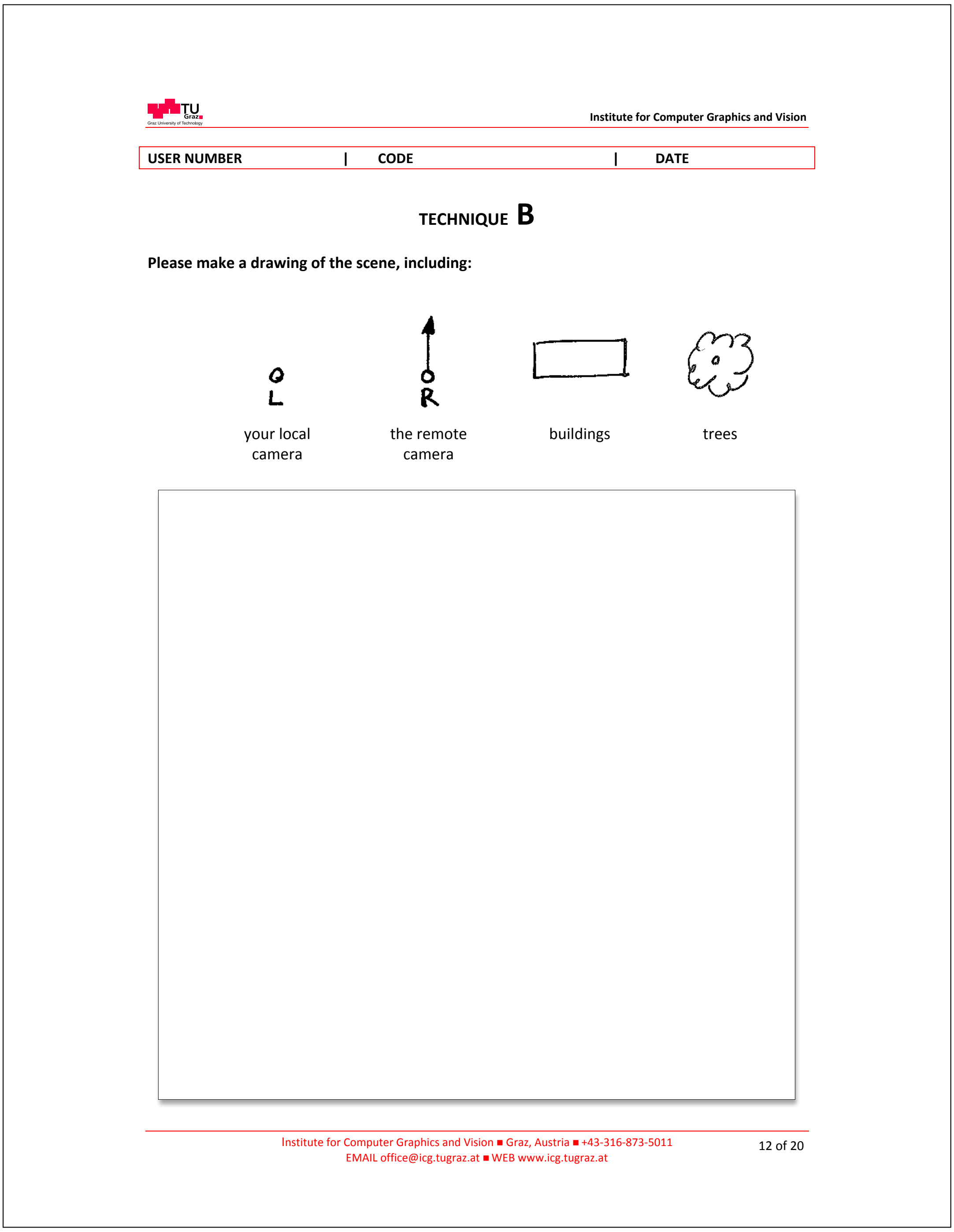
   I have seen it: **O** never **O** once **O** several times **O** daily

2. Can you see the remote camera?

   O yes, I can see the camera **O** no, I can not see the camera

3. From your location, can you see any object that is also visible in the remote camera?

   O yes **O** no **O** not sure



TU Graz — Graz University of Technology

Institute for Computer Graphics and Vision

USER NUMBER	CODE	DATE

TECHNIQUE **B**

Please make a drawing of the scene, including:

your local camera

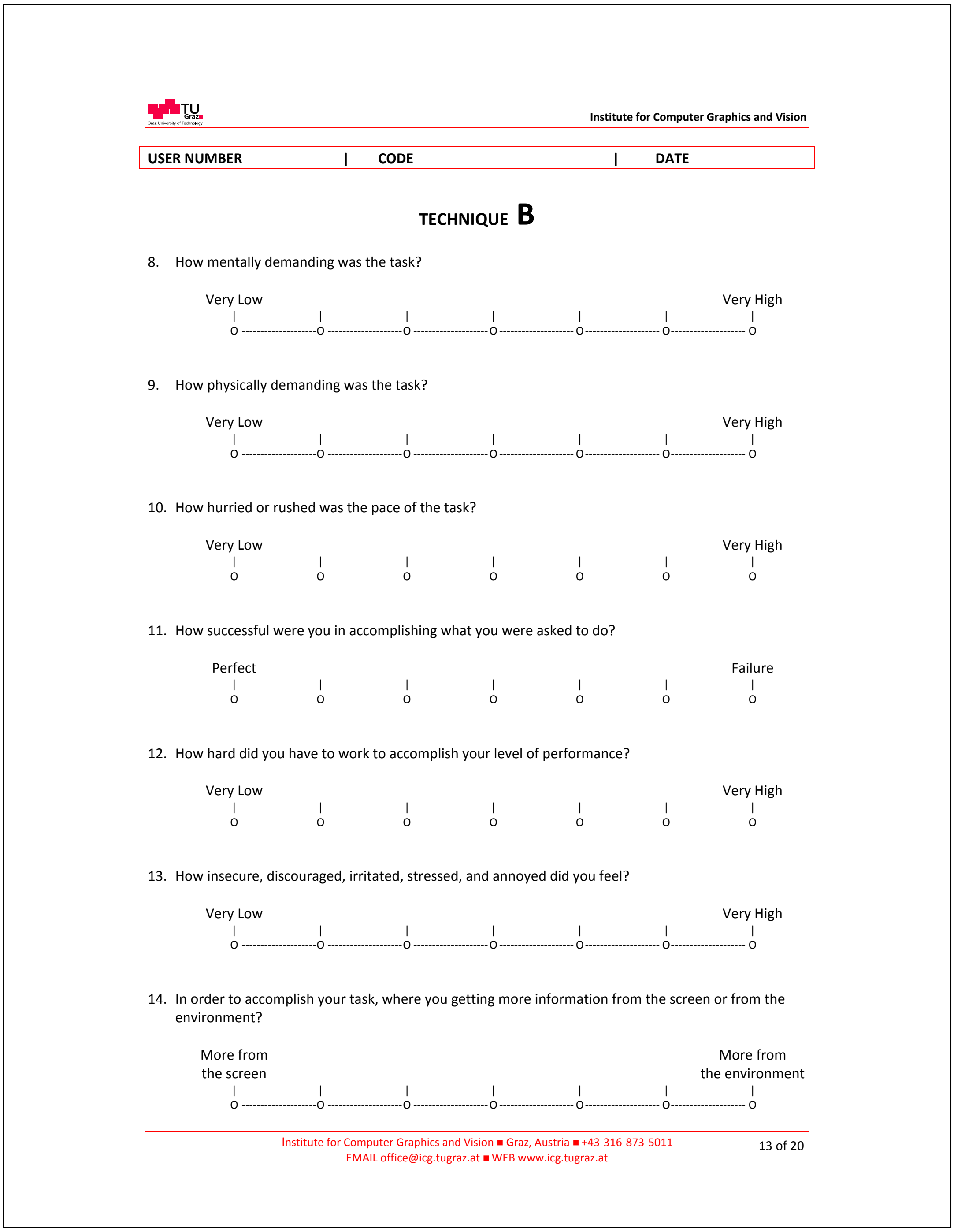
the remote camera

buildings

trees

Institute for Computer Graphics and Vision ■ Graz, Austria ■ +43-316-873-5011
EMAIL office@icg.tugraz.at ■ WEB www.icg.tugraz.at

12 of 20



TU Graz
Graz University of Technology

Institute for Computer Graphics and Vision

USER NUMBER | CODE | DATE

TECHNIQUE B

8. How mentally demanding was the task?

Very Low — Very High

```
|         |         |         |         |         |         |
O --------O --------O --------O --------O --------O --------O
```

9. How physically demanding was the task?

Very Low — Very High

```
|         |         |         |         |         |         |
O --------O --------O --------O --------O --------O --------O
```

10. How hurried or rushed was the pace of the task?

Very Low — Very High

```
|         |         |         |         |         |         |
O --------O --------O --------O --------O --------O --------O
```

11. How successful were you in accomplishing what you were asked to do?

Perfect — Failure

```
|         |         |         |         |         |         |
O --------O --------O --------O --------O --------O --------O
```

12. How hard did you have to work to accomplish your level of performance?

Very Low — Very High

```
|         |         |         |         |         |         |
O --------O --------O --------O --------O --------O --------O
```

13. How insecure, discouraged, irritated, stressed, and annoyed did you feel?

Very Low — Very High

```
|         |         |         |         |         |         |
O --------O --------O --------O --------O --------O --------O
```

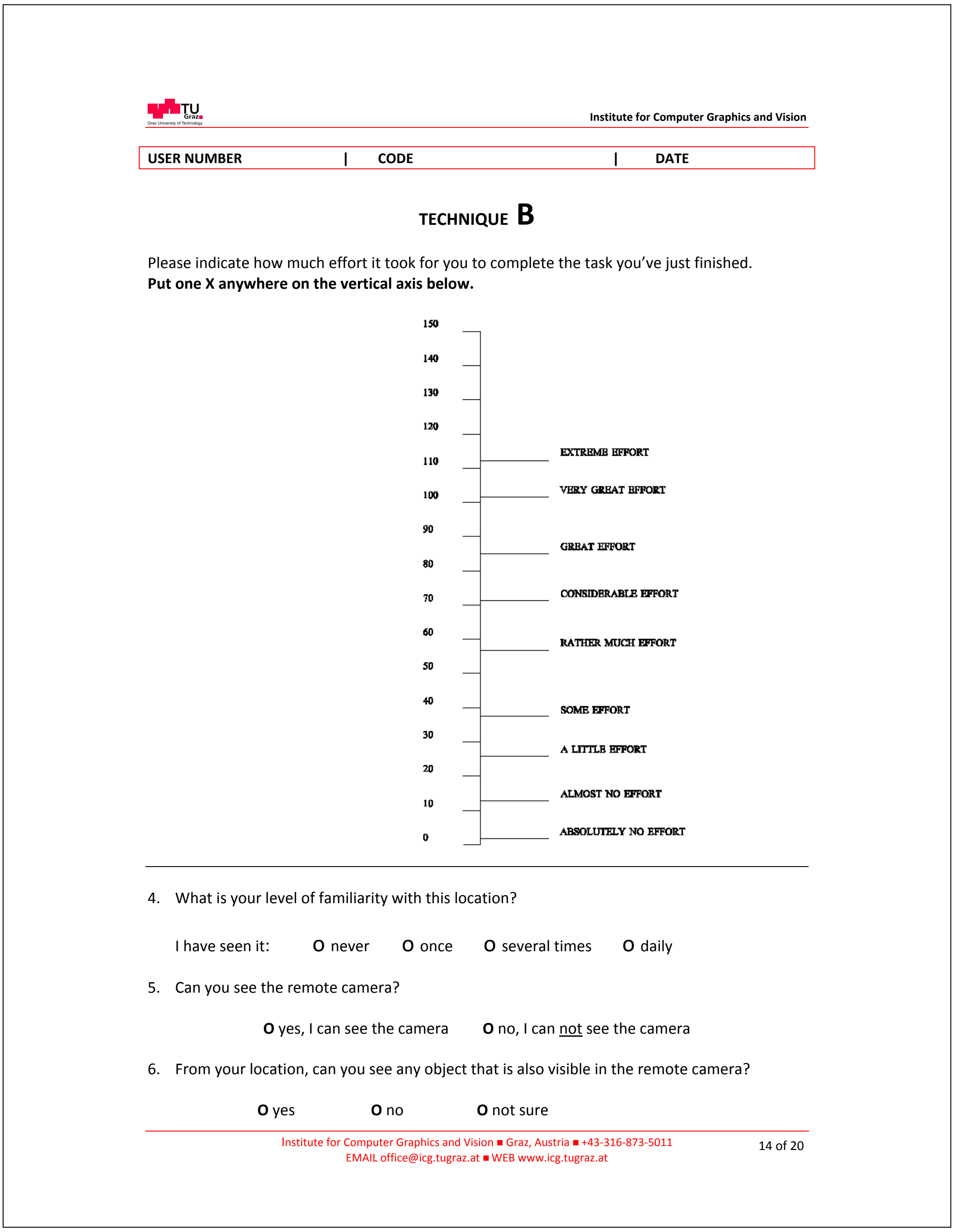
14. In order to accomplish your task, where you getting more information from the screen or from the environment?

More from the screen — More from the environment

```
|         |         |         |         |         |         |
O --------O --------O --------O --------O --------O --------O
```

Institute for Computer Graphics and Vision ■ Graz, Austria ■ +43-316-873-5011
EMAIL office@icg.tugraz.at ■ WEB www.icg.tugraz.at

13 of 20



TU Graz Graz University of Technology

Institute for Computer Graphics and Vision

USER NUMBER	CODE	DATE

TECHNIQUE B

Please indicate how much effort it took for you to complete the task you've just finished.
Put one X anywhere on the vertical axis below.

150
140
130
120
EXTREME EFFORT
110
VERY GREAT EFFORT
100
90
GREAT EFFORT
80
CONSIDERABLE EFFORT
70
60
RATHER MUCH EFFORT
50
40
SOME EFFORT
30
A LITTLE EFFORT
20
ALMOST NO EFFORT
10
ABSOLUTELY NO EFFORT
0

4. What is your level of familiarity with this location?

I have seen it: **O** never **O** once **O** several times **O** daily

5. Can you see the remote camera?

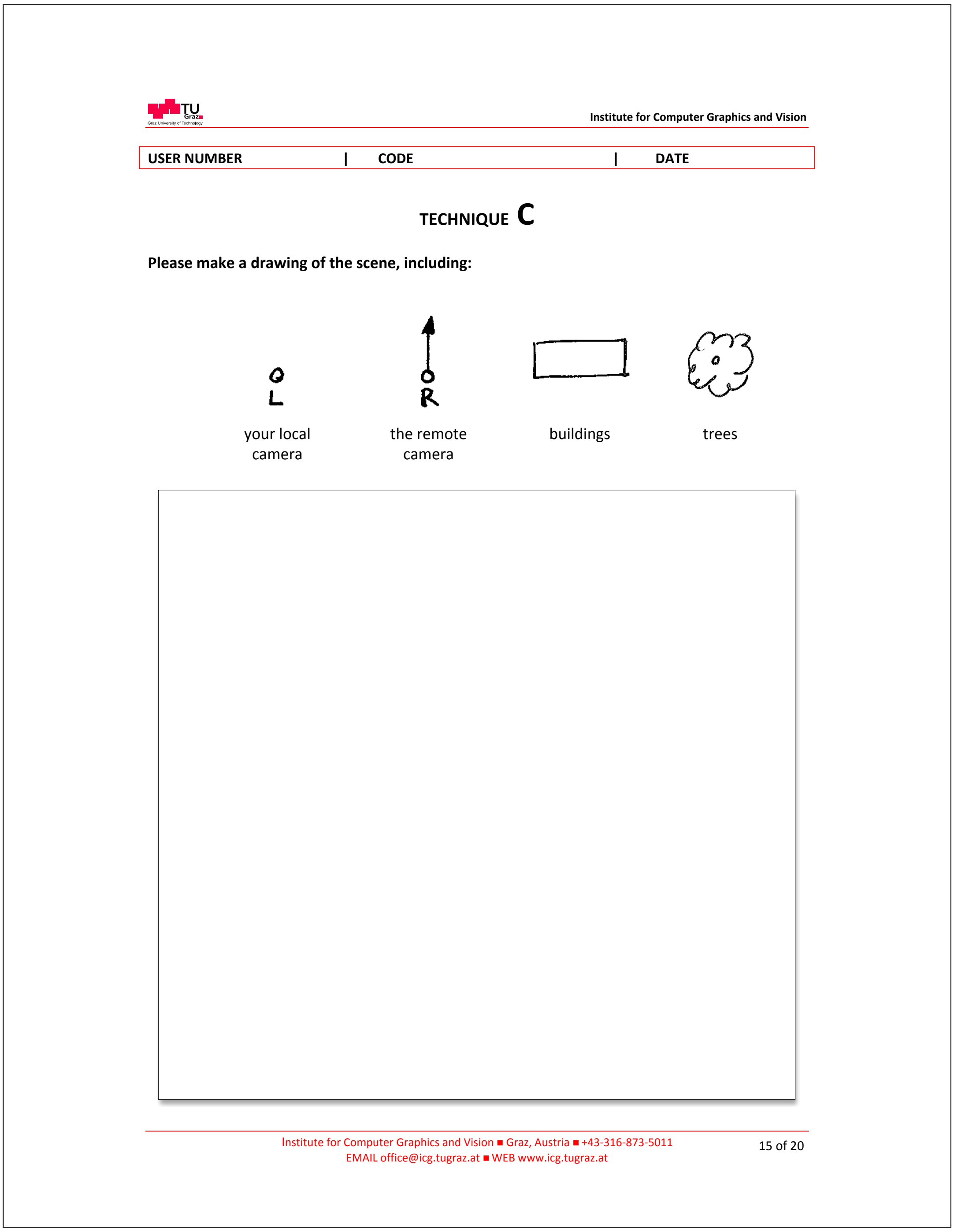
O yes, I can see the camera **O** no, I can not see the camera

6. From your location, can you see any object that is also visible in the remote camera?

O yes **O** no **O** not sure

Institute for Computer Graphics and Vision ■ Graz, Austria ■ +43-316-873-5011
EMAIL office@icg.tugraz.at ■ WEB www.icg.tugraz.at

14 of 20



Institute for Computer Graphics and Vision

USER NUMBER | CODE | DATE

TECHNIQUE C

Please make a drawing of the scene, including:

your local camera

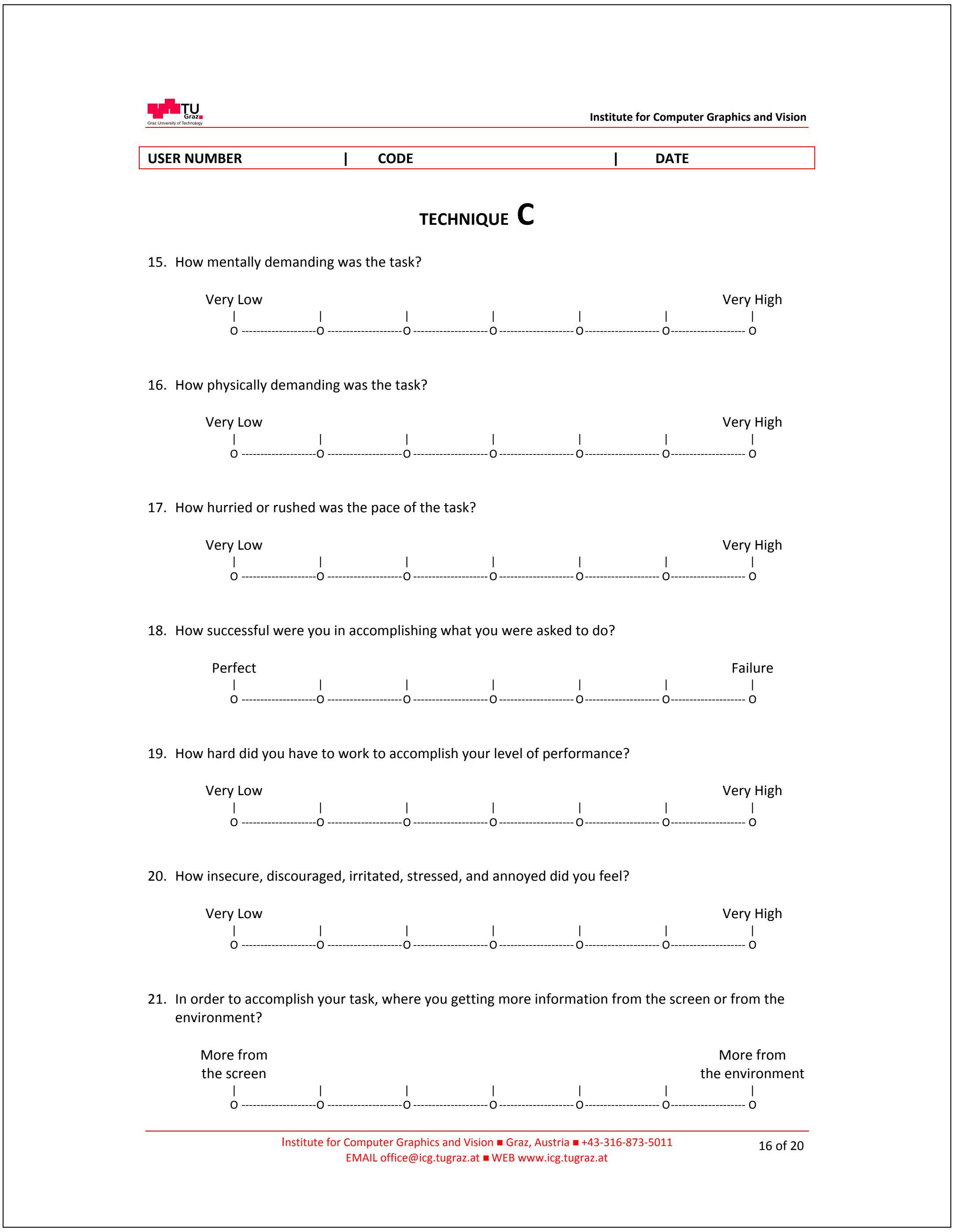
the remote camera

buildings

trees

Institute for Computer Graphics and Vision ■ Graz, Austria ■ +43-316-873-5011
EMAIL office@icg.tugraz.at ■ WEB www.icg.tugraz.at

15 of 20



Institute for Computer Graphics and Vision

USER NUMBER | CODE | DATE

TECHNIQUE C

15. How mentally demanding was the task?

```
Very Low                                                                Very High
   |            |            |            |            |            |            |
   O -----------O -----------O -----------O -----------O -----------O----------- O
```

16. How physically demanding was the task?

```
Very Low                                                                Very High
   |            |            |            |            |            |            |
   O -----------O -----------O -----------O -----------O -----------O----------- O
```

17. How hurried or rushed was the pace of the task?

```
Very Low                                                                Very High
   |            |            |            |            |            |            |
   O -----------O -----------O -----------O -----------O -----------O----------- O
```

18. How successful were you in accomplishing what you were asked to do?

```
 Perfect                                                                 Failure
   |            |            |            |            |            |            |
   O -----------O -----------O -----------O -----------O -----------O----------- O
```

19. How hard did you have to work to accomplish your level of performance?

```
Very Low                                                                Very High
   |            |            |            |            |            |            |
   O -----------O -----------O -----------O -----------O -----------O----------- O
```

20. How insecure, discouraged, irritated, stressed, and annoyed did you feel?

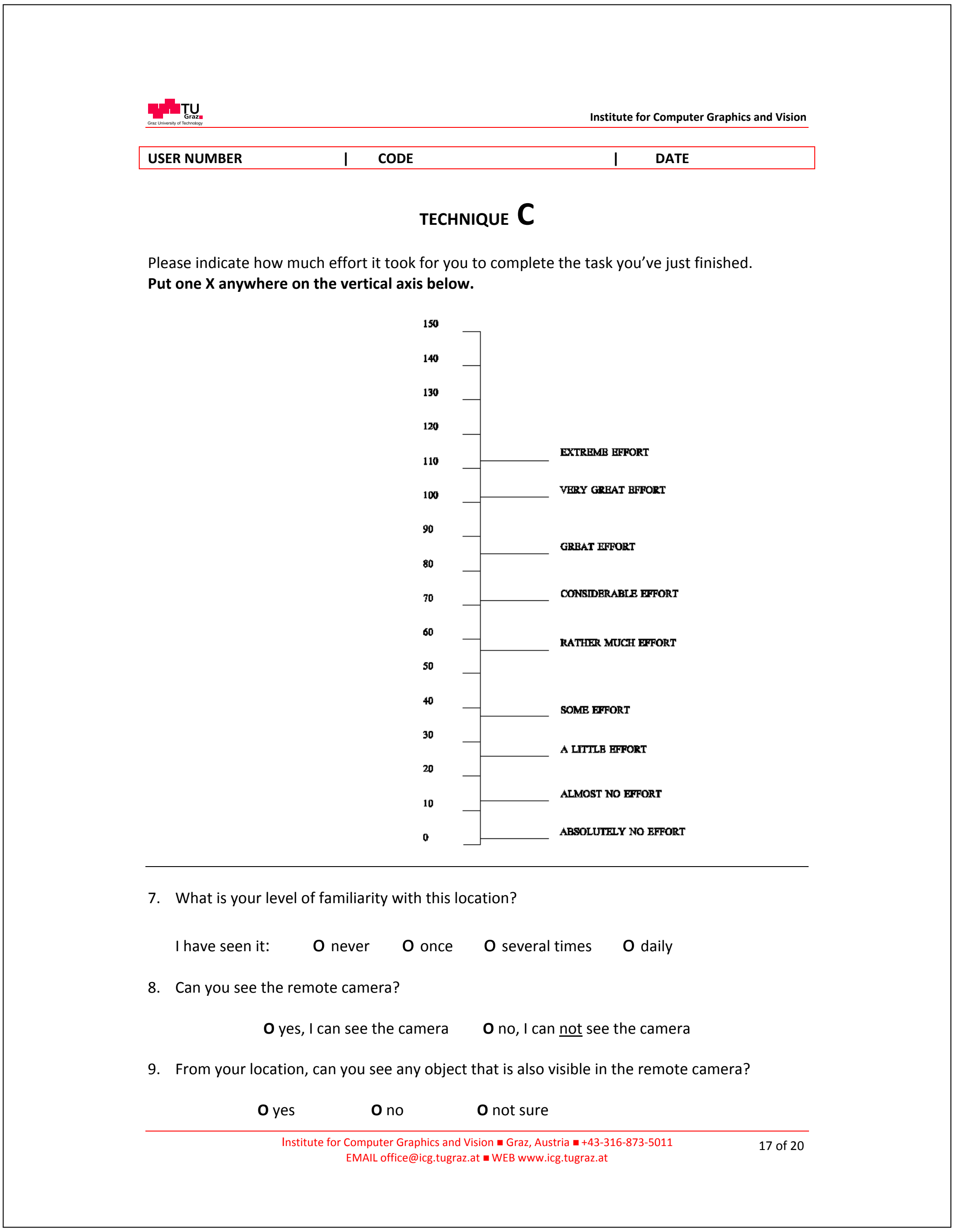
```
Very Low                                                                Very High
   |            |            |            |            |            |            |
   O -----------O -----------O -----------O -----------O -----------O----------- O
```

21. In order to accomplish your task, where you getting more information from the screen or from the environment?

```
More from                                                               More from
the screen                                                          the environment
   |            |            |            |            |            |            |
   O -----------O -----------O -----------O -----------O -----------O----------- O
```

Institute for Computer Graphics and Vision ■ Graz, Austria ■ +43-316-873-5011
EMAIL office@icg.tugraz.at ■ WEB www.icg.tugraz.at

16 of 20



USER NUMBER | CODE | DATE

TECHNIQUE **C**

Please indicate how much effort it took for you to complete the task you've just finished.
Put one X anywhere on the vertical axis below.

150
140
130
120
110 EXTREME EFFORT
100 VERY GREAT EFFORT
90
80 GREAT EFFORT
70 CONSIDERABLE EFFORT
60 RATHER MUCH EFFORT
50
40 SOME EFFORT
30 A LITTLE EFFORT
20
10 ALMOST NO EFFORT
0 ABSOLUTELY NO EFFORT

7. What is your level of familiarity with this location?

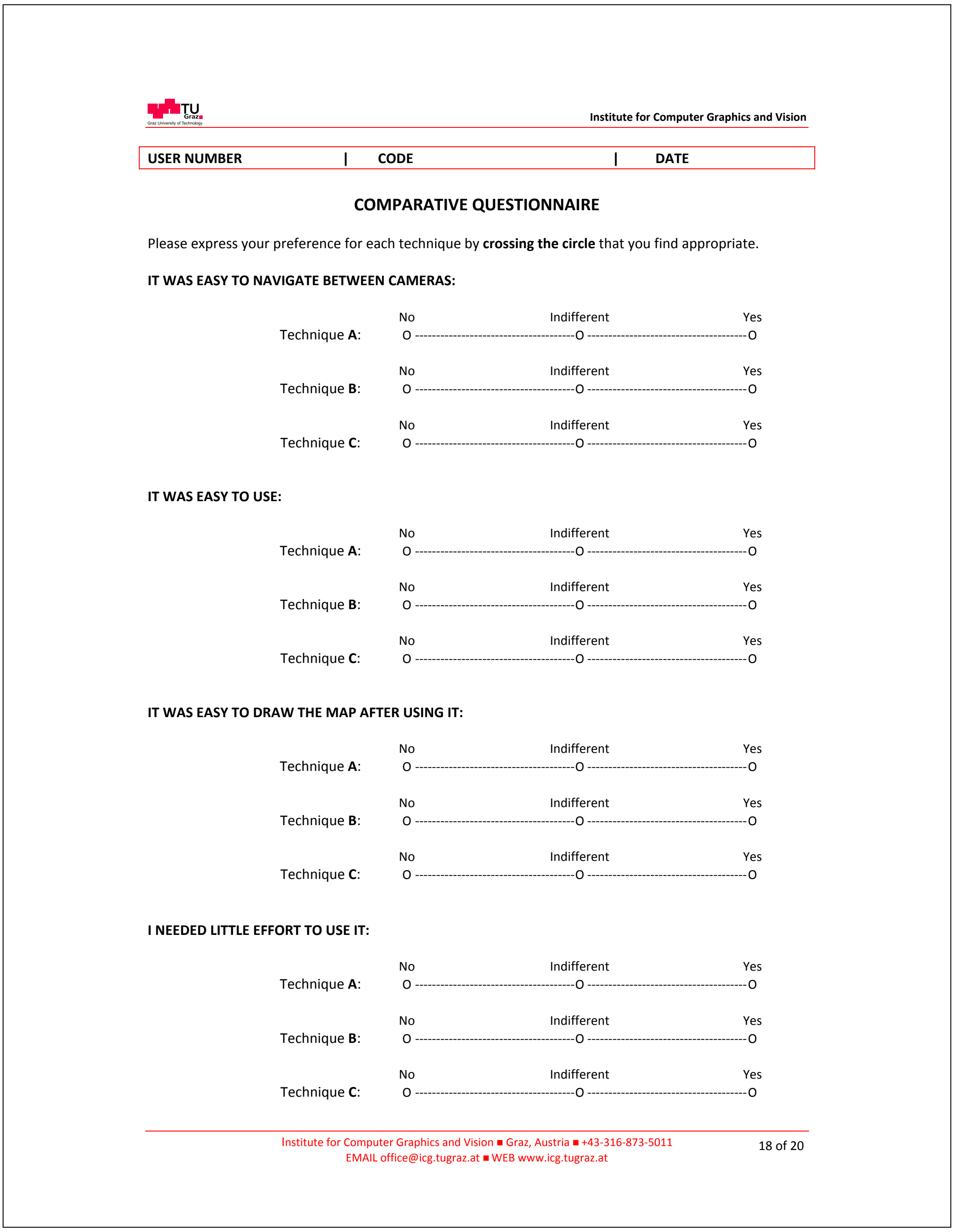
   I have seen it: **O** never **O** once **O** several times **O** daily

8. Can you see the remote camera?

   O yes, I can see the camera **O** no, I can not see the camera

9. From your location, can you see any object that is also visible in the remote camera?

   O yes **O** no **O** not sure



Institute for Computer Graphics and Vision

USER NUMBER | CODE | DATE

COMPARATIVE QUESTIONNAIRE

Please express your preference for each technique by **crossing the circle** that you find appropriate.

IT WAS EASY TO NAVIGATE BETWEEN CAMERAS:

No Indifferent Yes
Technique **A**: O -------------------------------------O -------------------------------------O

No Indifferent Yes
Technique **B**: O -------------------------------------O -------------------------------------O

No Indifferent Yes
Technique **C**: O -------------------------------------O -------------------------------------O

IT WAS EASY TO USE:

No Indifferent Yes
Technique **A**: O -------------------------------------O -------------------------------------O

No Indifferent Yes
Technique **B**: O -------------------------------------O -------------------------------------O

No Indifferent Yes
Technique **C**: O -------------------------------------O -------------------------------------O

IT WAS EASY TO DRAW THE MAP AFTER USING IT:

No Indifferent Yes
Technique **A**: O -------------------------------------O -------------------------------------O

No Indifferent Yes
Technique **B**: O -------------------------------------O -------------------------------------O

No Indifferent Yes
Technique **C**: O -------------------------------------O -------------------------------------O

I NEEDED LITTLE EFFORT TO USE IT:

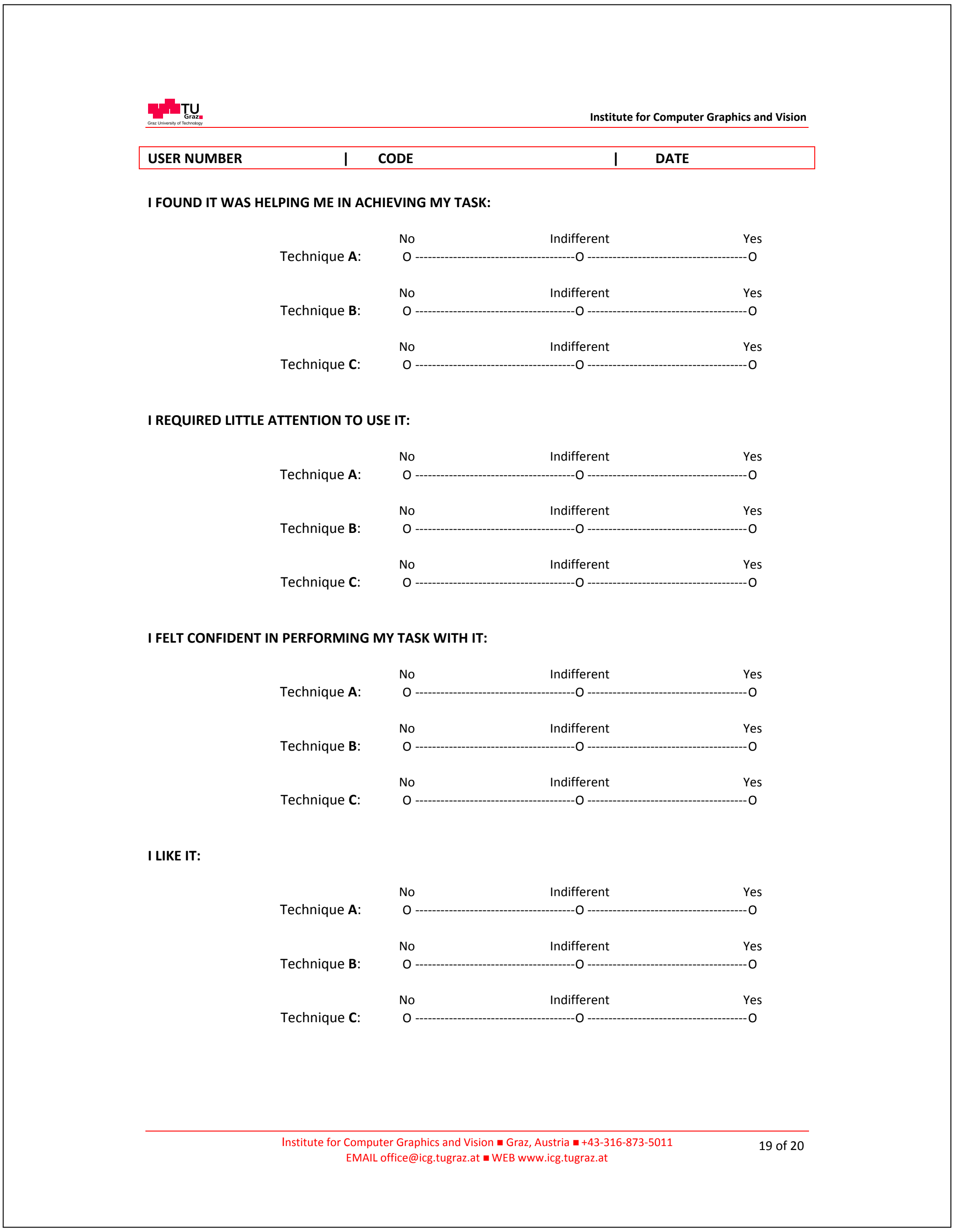
No Indifferent Yes
Technique **A**: O -------------------------------------O -------------------------------------O

No Indifferent Yes
Technique **B**: O -------------------------------------O -------------------------------------O

No Indifferent Yes
Technique **C**: O -------------------------------------O -------------------------------------O

Institute for Computer Graphics and Vision ■ Graz, Austria ■ +43-316-873-5011
EMAIL office@icg.tugraz.at ■ WEB www.icg.tugraz.at

18 of 20



Institute for Computer Graphics and Vision

USER NUMBER | CODE | DATE

I FOUND IT WAS HELPING ME IN ACHIEVING MY TASK:

Technique **A**: No O --------------------------------------O Indifferent --------------------------------------O Yes

Technique **B**: No O --------------------------------------O Indifferent --------------------------------------O Yes

Technique **C**: No O --------------------------------------O Indifferent --------------------------------------O Yes

I REQUIRED LITTLE ATTENTION TO USE IT:

Technique **A**: No O --------------------------------------O Indifferent --------------------------------------O Yes

Technique **B**: No O --------------------------------------O Indifferent --------------------------------------O Yes

Technique **C**: No O --------------------------------------O Indifferent --------------------------------------O Yes

I FELT CONFIDENT IN PERFORMING MY TASK WITH IT:

Technique **A**: No O --------------------------------------O Indifferent --------------------------------------O Yes

Technique **B**: No O --------------------------------------O Indifferent --------------------------------------O Yes

Technique **C**: No O --------------------------------------O Indifferent --------------------------------------O Yes

I LIKE IT:

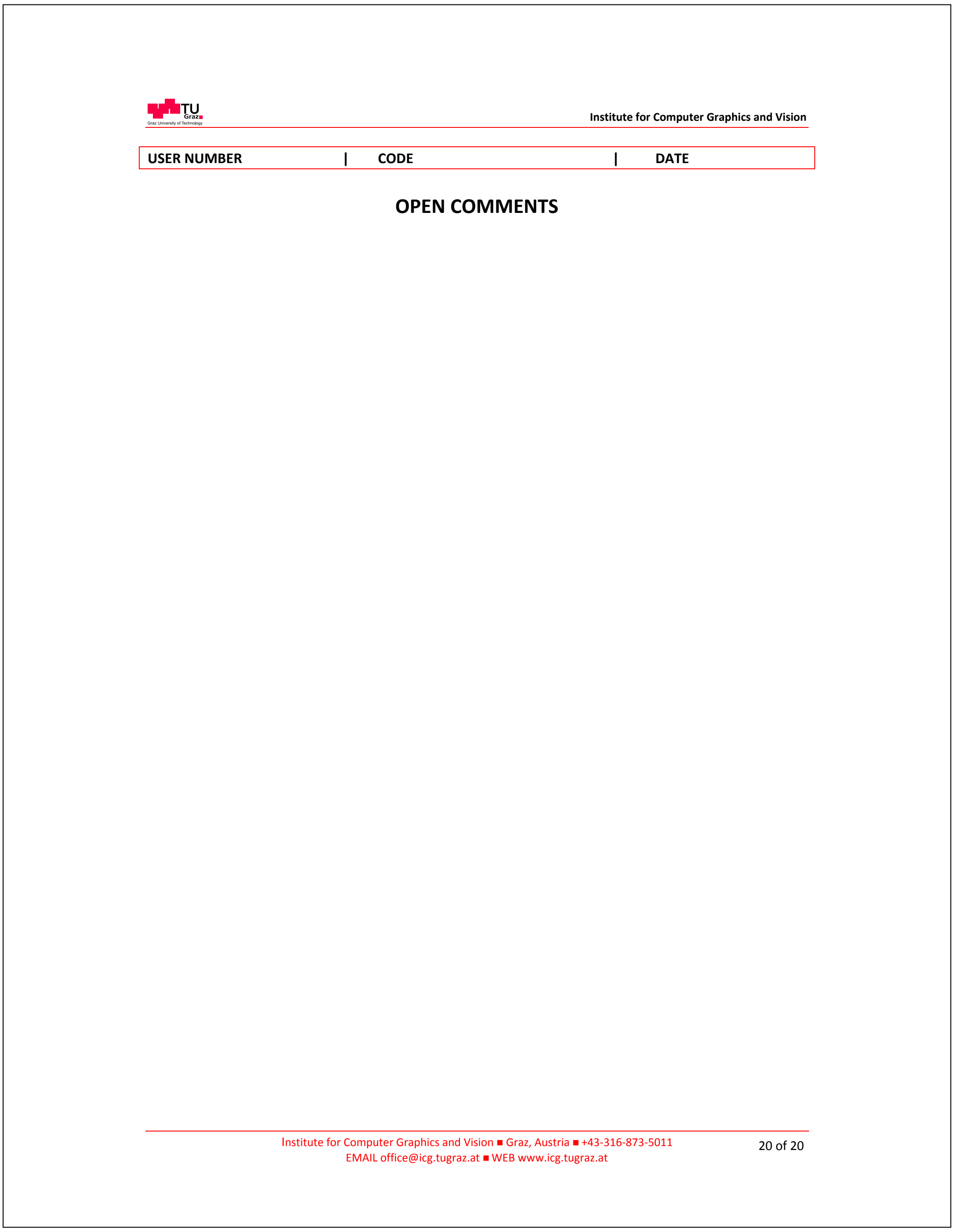
Technique **A**: No O --------------------------------------O Indifferent --------------------------------------O Yes

Technique **B**: No O --------------------------------------O Indifferent --------------------------------------O Yes

Technique **C**: No O --------------------------------------O Indifferent --------------------------------------O Yes

Institute for Computer Graphics and Vision ■ Graz, Austria ■ +43-316-873-5011
EMAIL office@icg.tugraz.at ■ WEB www.icg.tugraz.at

19 of 20



Institute for Computer Graphics and Vision

USER NUMBER | CODE | DATE

OPEN COMMENTS

Institute for Computer Graphics and Vision ■ Graz, Austria ■ +43-316-873-5011
EMAIL office@icg.tugraz.at ■ WEB www.icg.tugraz.at

20 of 20

A.2 Pointing to multiple augmentations (pilot)

This section presents the information sheet and the questionnaire used in the pilot study presented in Section 5.2.2.

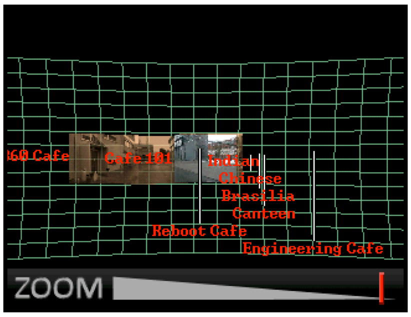
The study

You will try three different user interfaces that allow you to browse spatial information. In this particular case, the interfaces will allow you to browse all cafes and restaurants in the University campus. All techniques will present labels with the names of the cafes in Augmented Reality (AR).

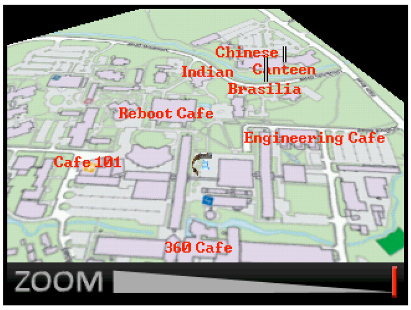
You are not assigned a particular task: we just ask you to spend a few minutes trying all the three interfaces. We are looking for preliminary feedback on our interfaces, so we will ask you to fill in a short questionnaire with informal feedback afterwards.

At the moment no occlusion clues are implemented, so cafes that are behind a building will also be visible in your AR view. The system is still in an early stage, so please be forgiving if the tracking is not as robust as it could (should) be!

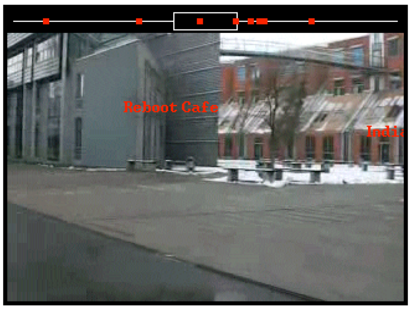
The techniques



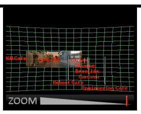
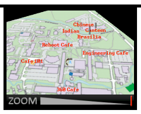
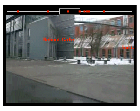
Zooming interface. This interface allows you to zoom out the view, up to a virtual wide-angle view covering 360°. Your camera leaves a trace in the form of a panorama, to hint at the areas you already explored. A green wireframe grid shows you the extents of the space.

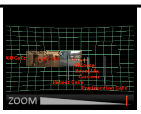
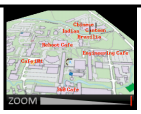


Transitional interface. This interface allows you to transition to a third-person view. This view shows you a map of the campus, with labels presenting the location of all cafes. Your position and live video frame is always in the middle of the screen, and the map rotates as you turn your camera.



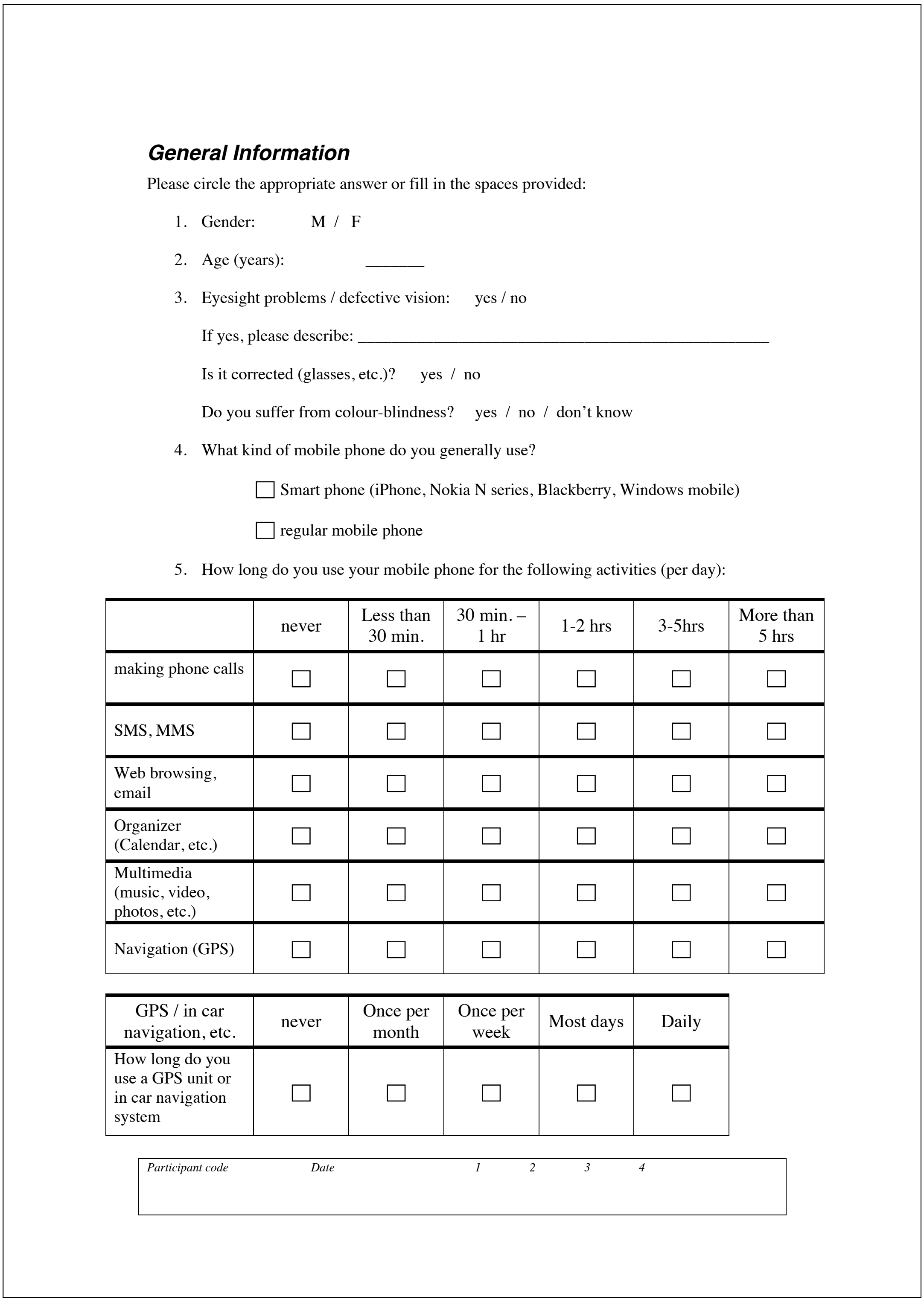
Compass overlay. This technique uses an overlay (on the upper side of the screen) to present as red dots all cafes/restaurants that are outside the current view of the camera (on the line) and inside the current view (represented by the square in the middle of the overlay).

	 Zooming interface	 Transitional interface
What did you like of it, and what didn't you like? What would you change?		
With respect to the *compass overlay* what did you like more and what less? 		

	 Zooming interface	 Transitional interface
Which interface would you prefer if the cafe you are looking for were not visible because it's behind another building?		
Which interface would you prefer if the cafe you are looking for were visible in the real world?		
Would you prefer having a user interface that combines both this technique and the compass overlay at the same time?		

A.3 Pointing to multiple augmentations

This section presents the questionnaire used in the user study presented in Section 5.2.4.



General Information

Please circle the appropriate answer or fill in the spaces provided:

1. Gender: M / F
2. Age (years): _______
3. Eyesight problems / defective vision: yes / no

   If yes, please describe: ____

   Is it corrected (glasses, etc.)? yes / no

   Do you suffer from colour-blindness? yes / no / don't know
4. What kind of mobile phone do you generally use?

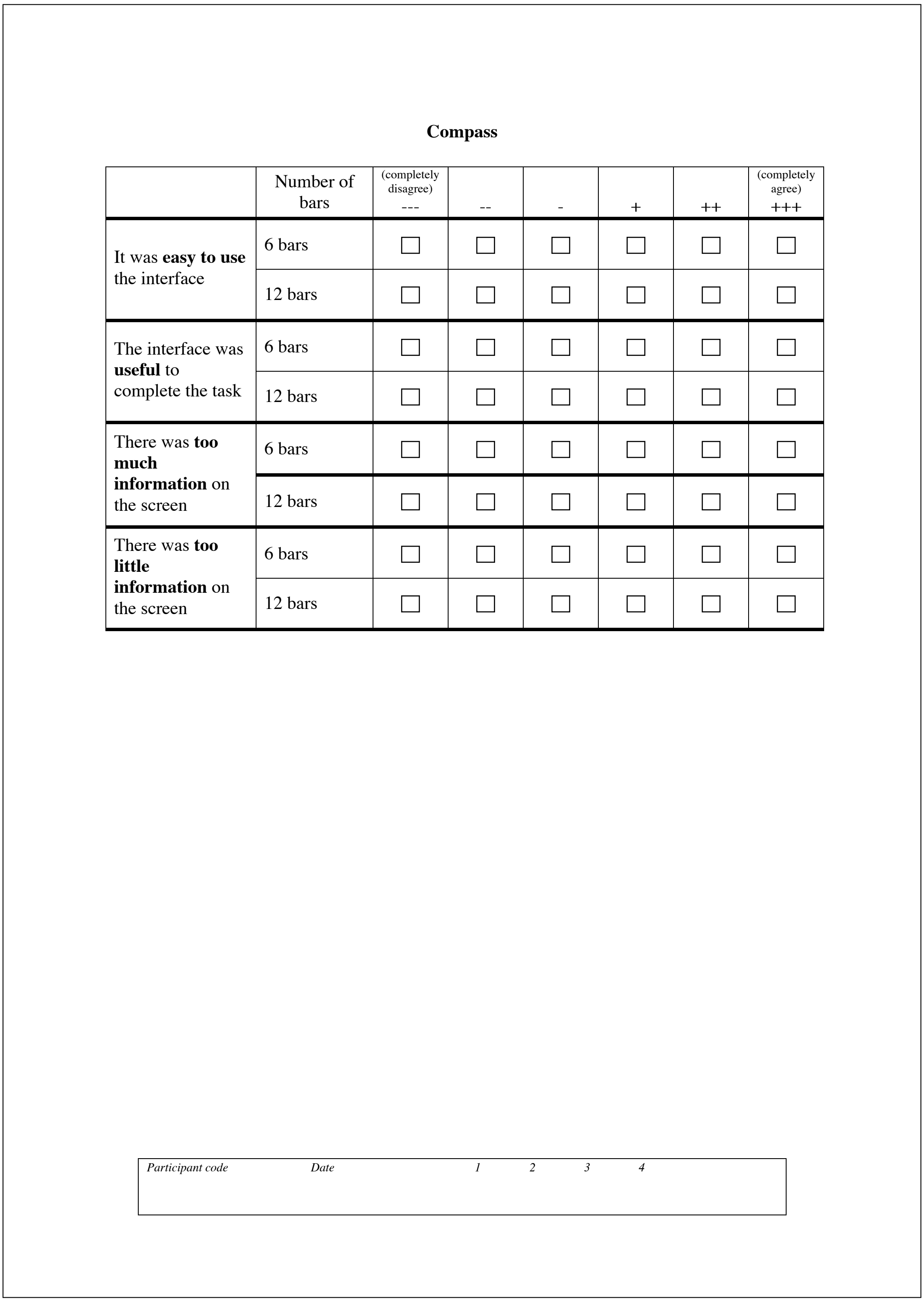
   □ Smart phone (iPhone, Nokia N series, Blackberry, Windows mobile)

   □ regular mobile phone
5. How long do you use your mobile phone for the following activities (per day):

	never	Less than 30 min.	30 min. – 1 hr	1-2 hrs	3-5hrs	More than 5 hrs
making phone calls	□	□	□	□	□	□
SMS, MMS	□	□	□	□	□	□
Web browsing, email	□	□	□	□	□	□
Organizer (Calendar, etc.)	□	□	□	□	□	□
Multimedia (music, video, photos, etc.)	□	□	□	□	□	□
Navigation (GPS)	□	□	□	□	□	□

GPS / in car navigation, etc.	never	Once per month	Once per week	Most days	Daily
How long do you use a GPS unit or in car navigation system	□	□	□	□	□

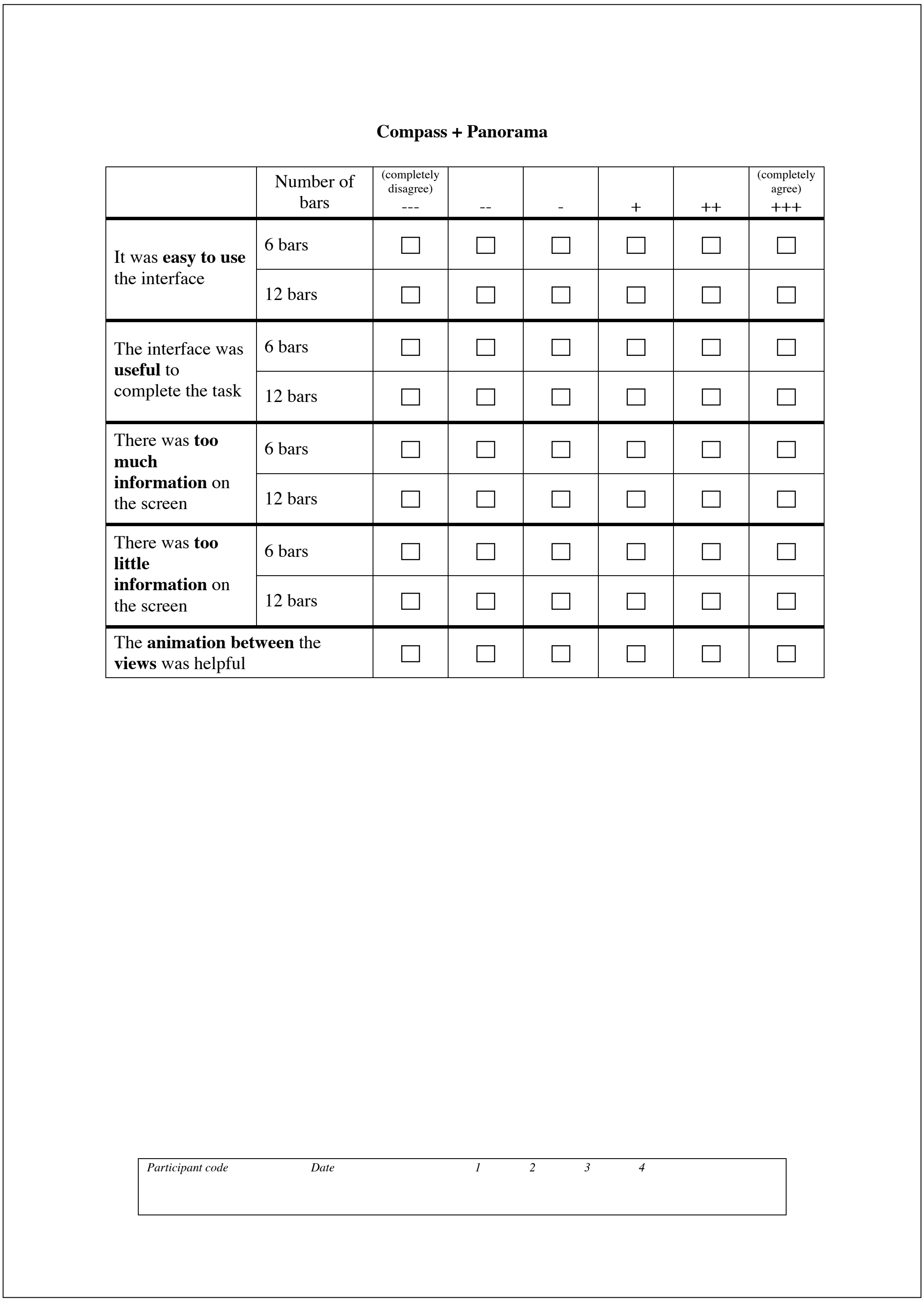
Participant code *Date* *1* *2* *3* *4*



Compass

	Number of bars	(completely disagree) ---	--	-	+	++	(completely agree) +++
It was **easy to use** the interface	6 bars	□	□	□	□	□	□
	12 bars	□	□	□	□	□	□
The interface was **useful** to complete the task	6 bars	□	□	□	□	□	□
	12 bars	□	□	□	□	□	□
There was **too much information** on the screen	6 bars	□	□	□	□	□	□
	12 bars	□	□	□	□	□	□
There was **too little information** on the screen	6 bars	□	□	□	□	□	□
	12 bars	□	□	□	□	□	□

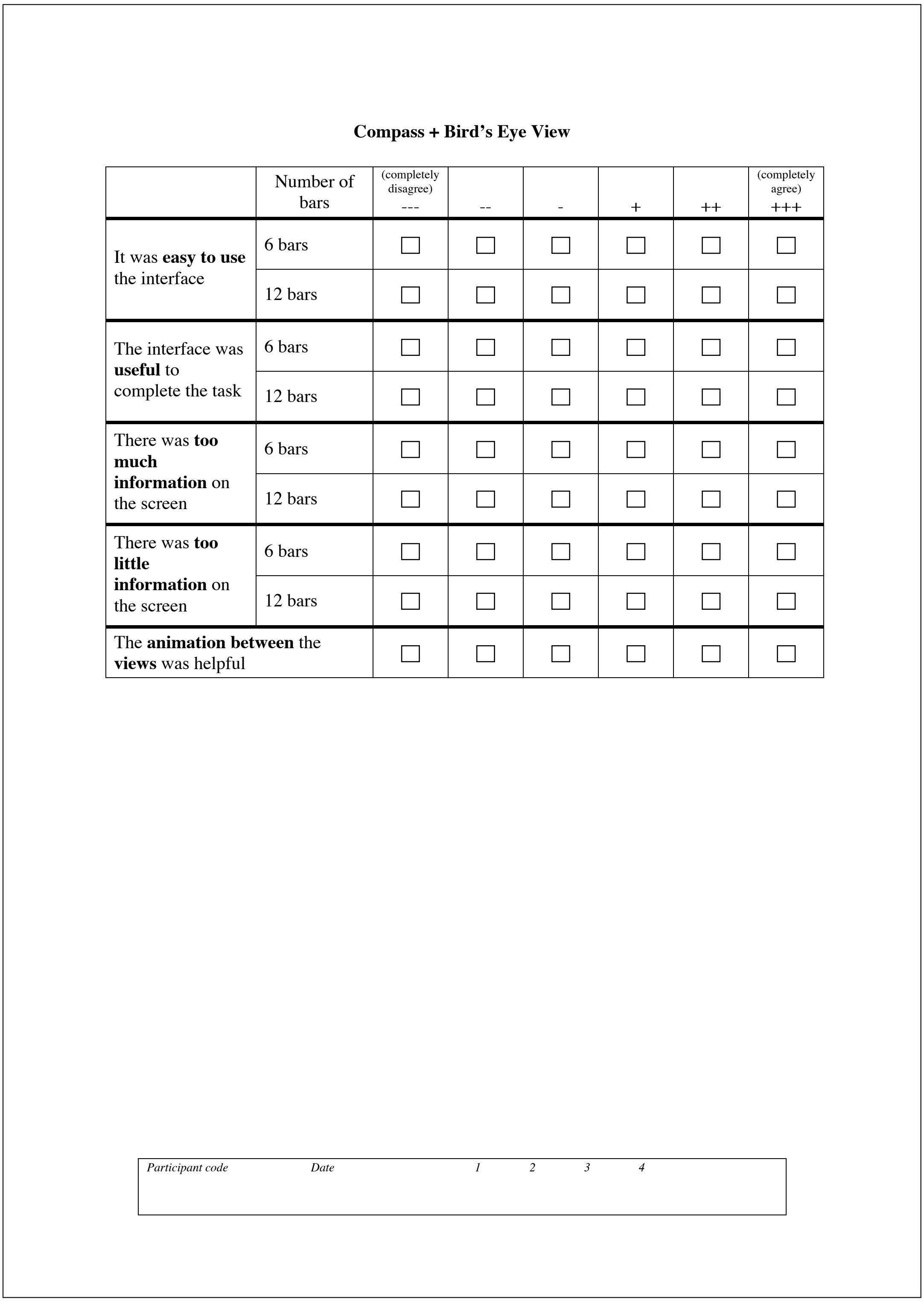
Participant code *Date* *1* *2* *3* *4*



Compass + Panorama

	Number of bars	(completely disagree) ---	--	-	+	++	(completely agree) +++
It was **easy to use** the interface	6 bars	□	□	□	□	□	□
	12 bars	□	□	□	□	□	□
The interface was **useful** to complete the task	6 bars	□	□	□	□	□	□
	12 bars	□	□	□	□	□	□
There was **too much information** on the screen	6 bars	□	□	□	□	□	□
	12 bars	□	□	□	□	□	□
There was **too little information** on the screen	6 bars	□	□	□	□	□	□
	12 bars	□	□	□	□	□	□
The **animation between** the **views** was helpful		□	□	□	□	□	□

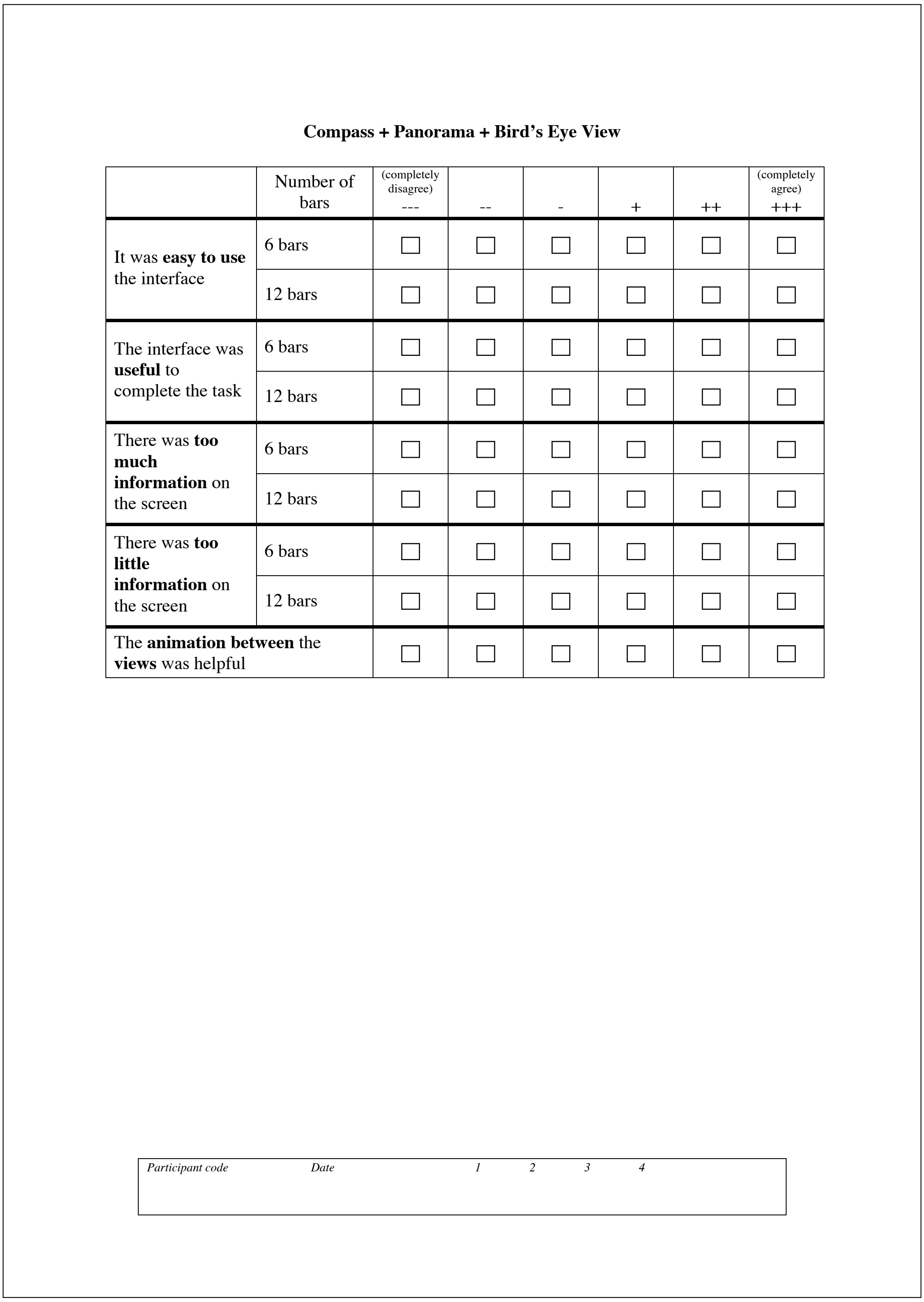
Participant code *Date* *1* *2* *3* *4*



Compass + Bird's Eye View

	Number of bars	(completely disagree) ---	--	-	+	++	(completely agree) +++
It was **easy to use** the interface	6 bars	□	□	□	□	□	□
	12 bars	□	□	□	□	□	□
The interface was **useful** to complete the task	6 bars	□	□	□	□	□	□
	12 bars	□	□	□	□	□	□
There was **too much information** on the screen	6 bars	□	□	□	□	□	□
	12 bars	□	□	□	□	□	□
There was **too little information** on the screen	6 bars	□	□	□	□	□	□
	12 bars	□	□	□	□	□	□
The **animation between** the **views** was helpful		□	□	□	□	□	□

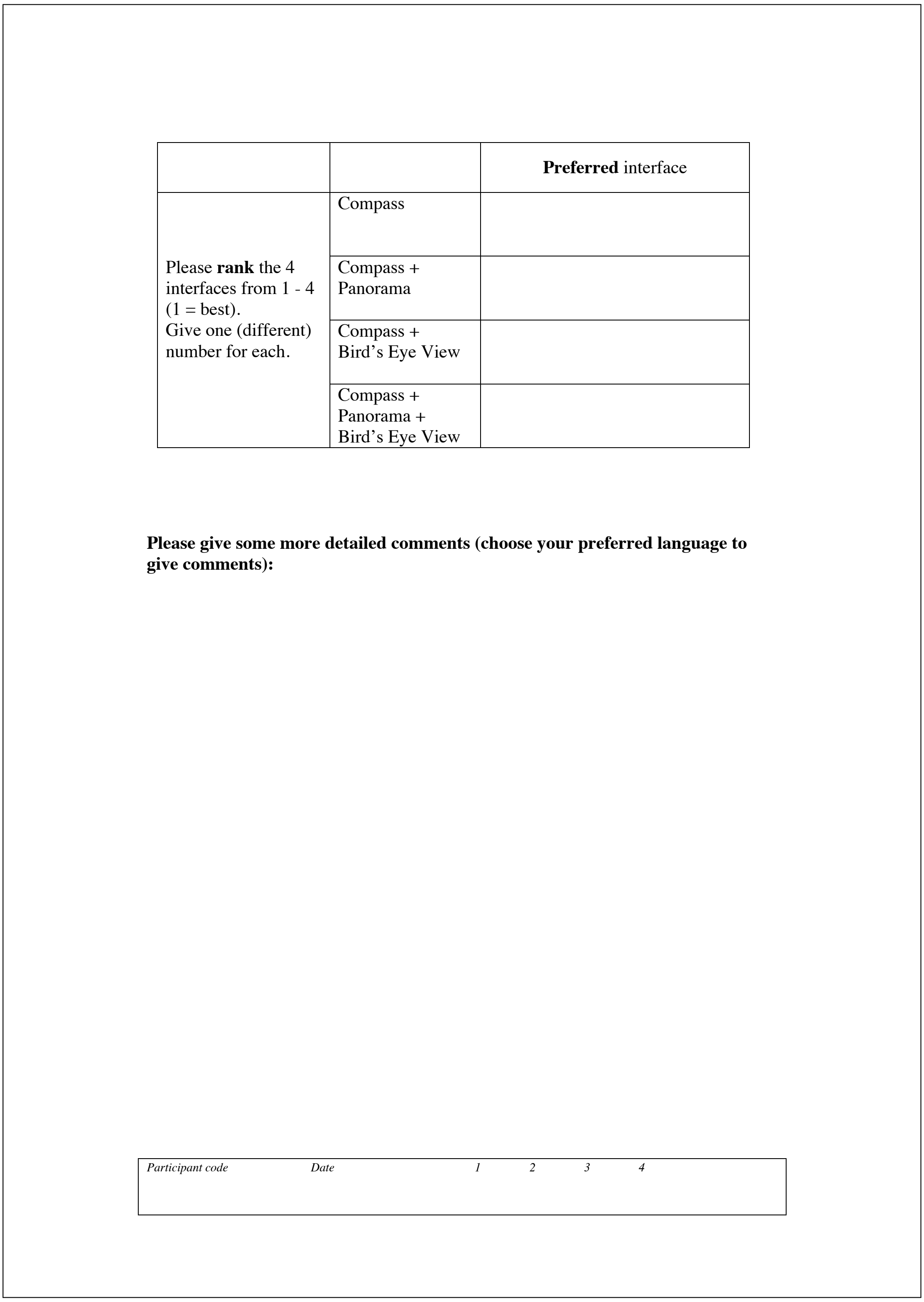
Participant code *Date* *1* *2* *3* *4*



Compass + Panorama + Bird's Eye View

	Number of bars	(completely disagree) ---	--	-	+	++	(completely agree) +++
It was **easy to use** the interface	6 bars	□	□	□	□	□	□
	12 bars	□	□	□	□	□	□
The interface was **useful** to complete the task	6 bars	□	□	□	□	□	□
	12 bars	□	□	□	□	□	□
There was **too much information** on the screen	6 bars	□	□	□	□	□	□
	12 bars	□	□	□	□	□	□
There was **too little information** on the screen	6 bars	□	□	□	□	□	□
	12 bars	□	□	□	□	□	□
The **animation between** the **views** was helpful		□	□	□	□	□	□

Participant code *Date* *1* *2* *3* *4*



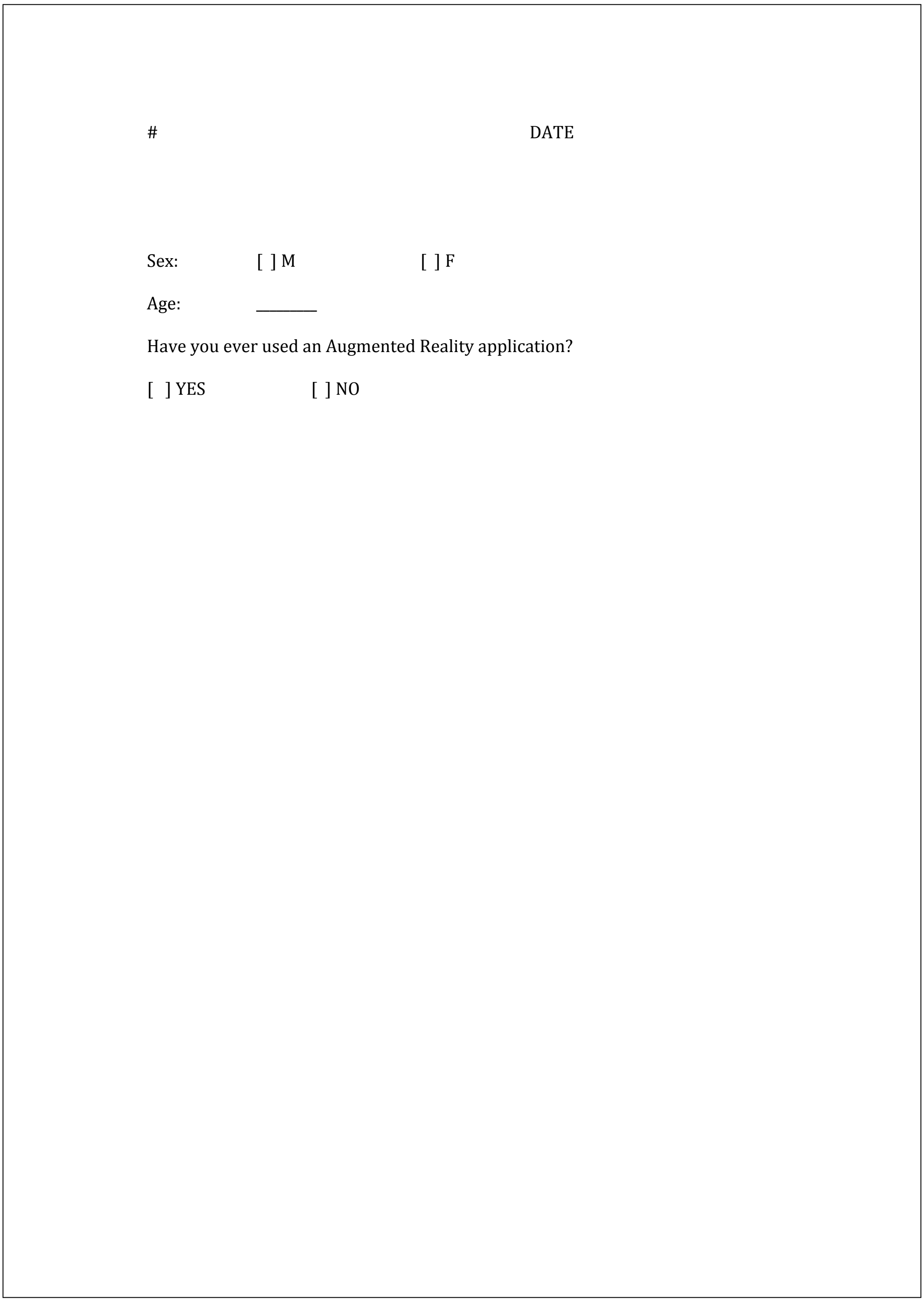
		Preferred interface
Please **rank** the 4 interfaces from 1 - 4 (1 = best). Give one (different) number for each.	Compass	
	Compass + Panorama	
	Compass + Bird’s Eye View	
	Compass + Panorama + Bird’s Eye View	

Please give some more detailed comments (choose your preferred language to give comments):

Participant code	*Date*	*1*	*2*	*3*	*4*

A.4 Panoramic overviews (pilot)

This section presents the information sheet and the questionnaire used in the pilot study presented in Section 5.3.2.



#   DATE

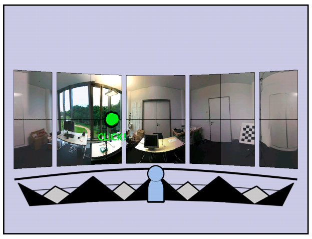
Sex: [ ] M   [ ] F

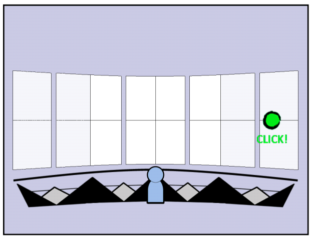
Age: ________

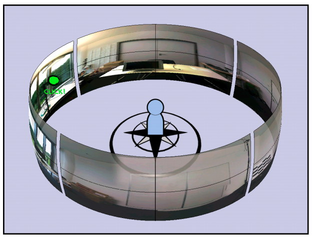
Have you ever used an Augmented Reality application?

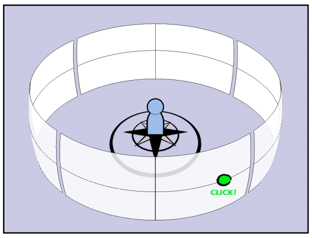
[ ] YES   [ ] NO

What was your **strategy** for resolving the physical direction in the following different cases?



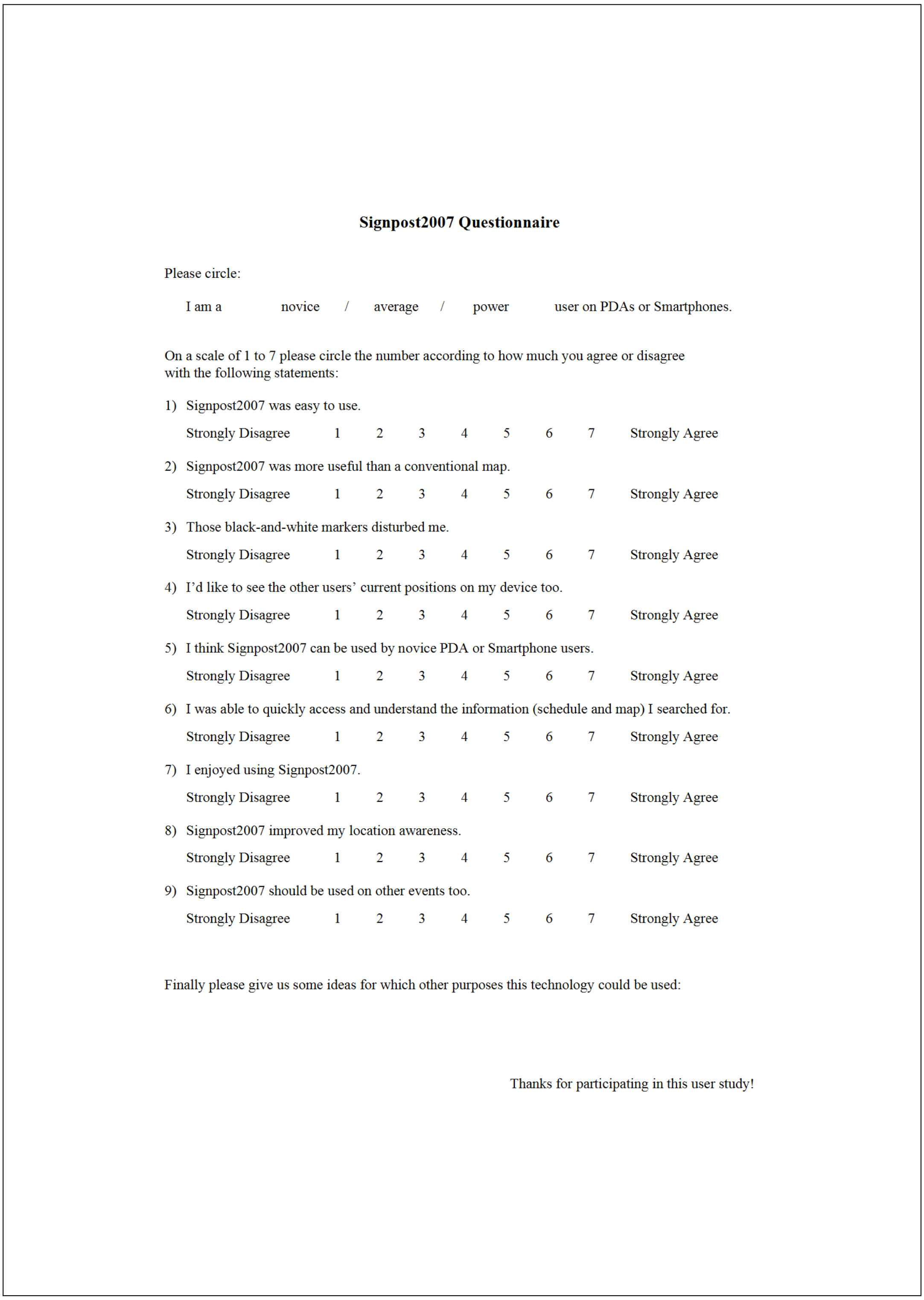






A.5 Indoor navigation (MEDC 2007)

This section presents the questionnaire used at the Microsoft MEDC 2007 conference, where our system for indoor navigation, presented in Section 8.1.3, was deployed. Please note that the MEDC 2007 study is work that was done prior to this thesis, and the author of this thesis did not contribute to it. This study is mostly work done by Daniel Wagner (TU Graz) and István Barakonyi (Imagination GmbH).



Signpost2007 Questionnaire

Please circle:

I am a novice / average / power user on PDAs or Smartphones.

On a scale of 1 to 7 please circle the number according to how much you agree or disagree with the following statements:

1) Signpost2007 was easy to use.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

2) Signpost2007 was more useful than a conventional map.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

3) Those black-and-white markers disturbed me.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

4) I'd like to see the other users' current positions on my device too.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

5) I think Signpost2007 can be used by novice PDA or Smartphone users.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

6) I was able to quickly access and understand the information (schedule and map) I searched for.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

7) I enjoyed using Signpost2007.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

8) Signpost2007 improved my location awareness.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

9) Signpost2007 should be used on other events too.

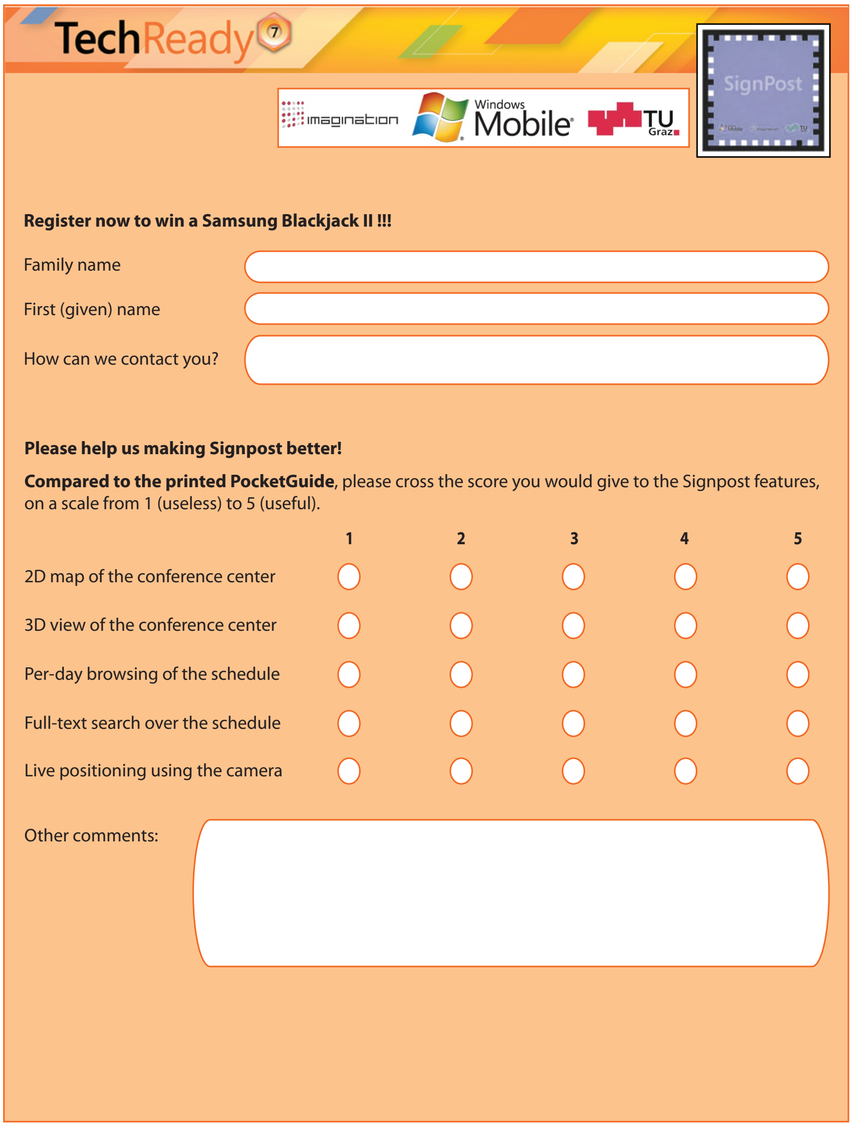
Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

Finally please give us some ideas for which other purposes this technology could be used:

Thanks for participating in this user study!

A.6 Indoor navigation (TechReady 7)

This section presents the questionnaire used at the Microsoft TechReady 7 conference, where we deployed our system for indoor navigation presented in Section 8.1.3.


TechReady 7
imagination
Windows Mobile
TU Graz
SignPost

Register now to win a Samsung Blackjack II !!!

Family name

First (given) name

How can we contact you?

Please help us making Signpost better!

Compared to the printed PocketGuide, please cross the score you would give to the Signpost features, on a scale from 1 (useless) to 5 (useful).

	1	2	3	4	5
2D map of the conference center	○	○	○	○	○
3D view of the conference center	○	○	○	○	○
Per-day browsing of the schedule	○	○	○	○	○
Full-text search over the schedule	○	○	○	○	○
Live positioning using the camera	○	○	○	○	○

Other comments:

Appendix B
Acronyms

AR	Augmented Reality
DOF	degrees of freedom
DoG	Difference of Gaussians
DoP	Dilution of Precision
EKF	extended Kalman filter
GPS	Global Positioning System
HCI	Human-Computer Interaction
HMDB	Hypermedia Database
M	Mean
MR	Mixed Reality
NCC	Normalised Cross-Correlation
RLE	run-length encoding
SD	Standard Deviation
SE	Standard Error
SIFT	Scale Invariant Feature Transform
SLAM	Simultaneous Localisation and Mapping
SSD	Sum of Squared Difference

SUS	System Usability Scale
VR	Virtual Reality
WIM	World-in-Miniature

Bibliography

[1] Gregory D. Abowd, Christopher G. Atkeson, Jason Hong, Sue Long, Rob Kooper, and Mike Pinkerton. *Cyberguide: a mobile context-aware tour guide.* Wireless Networks, 3:421–433, October 1997.

[2] Gary L. Allen. *Principles and practices for communicating route knowledge.* Applied Cognitive Psychology, 14(4):333–359, July 2000.

[3] Clemens Arth, Manfred Klopschitz, Gerhard Reitmayr, and Dieter Schmalstieg. *Real-time self-localization from panoramic images on mobile devices.* In Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality, ISMAR '11, page 37–46, Washington, DC, USA, 2011. IEEE Computer Society.

[4] Clemens Arth, Daniel Wagner, Manfred Klopschitz, Arnold Irschara, and Dieter Schmalstieg. *Wide area localization on mobile phones.* In IEEE / ACM International Symposium on Mixed and Augmented Reality, volume 0, pages 73–82, Los Alamitos, CA, USA, 2009. IEEE Computer Society.

[5] Carmen E. Au, Victor Ng, and James J. Clark. *Mirrormap: augmenting 2d mobile maps with virtual mirrors.* In Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, MobileHCI '11, page 255–264, New York, NY, USA, 2011. ACM.

[6] Benjamin Avery, Christian Sandor, and Bruce H. Thomas. *Improving Spatial Perception for Augmented Reality X-Ray Vision.* In Proceedings of the IEEE Virtual Reality Conference, pages 79–82. IEEE, March 2009.

[7] Benjamin Avery, Bruce H. Thomas, and Wayne Piekarski. *User evaluation of see-through vision for mobile outdoor augmented reality.* In Proceedings of the 7th

IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR '08, page 69–72, Washington, DC, USA, 2008. IEEE Computer Society.

[8] Ronald Azuma. *Tracking requirements for augmented reality.* Commun. ACM, 36(7):50–51, July 1993.

[9] Ryan Bane and Tobias Höllerer. *Interactive Tools for Virtual X-Ray Vision in Mobile Augmented Reality.* In Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR '04, page 231–239, Washington, DC, USA, 2004. IEEE Computer Society.

[10] Jörg Baus, Keith Cheverst, and Christian Kray. *A Survey of Map-based Mobile Guides.* In Map-based mobile services: theories, methods and implementations, pages 197–216. Springer, 2005.

[11] Ashweeni Kumar Beeharee and Anthony Steed. *A natural wayfinding exploiting photos in pedestrian navigation systems.* In Proceedings of the 8th conference on Human-computer interaction with mobile devices and services, MobileHCI '06, page 81–88, New York, NY, USA, 2006. ACM.

[12] Blaine Bell, Tobias Höllerer, and Steven Feiner. *An annotated situation-awareness aid for augmented reality.* In Proceedings of the 15th annual ACM symposium on User interface software and technology, UIST '02, page 213–216, New York, NY, USA, 2002. ACM.

[13] Dominik Bial, Dagmar Kern, Florian Alt, and Albrecht Schmidt. *Enhancing outdoor navigation systems through vibrotactile feedback.* In Proceedings of the 2011 annual conference on Human factors in computing systems (extended abstracts), CHI EA '11, page 1273–1278, New York, NY, USA, 2011. ACM.

[14] Frank Biocca, Arthur Tang, Charles Owen, and Fan Xiao. *Attention funnel: omnidirectional 3D cursor for mobile augmented reality platforms.* In Proceedings of the SIGCHI conference on Human Factors in computing systems, CHI '06, page 1115–1122, New York, NY, USA, 2006. ACM.

[15] Joachim Bobrich and Steffen Otto. *Augmented Maps.* In Geospatial Theory, Processing and Applications, volume IAPRS 34 of 4, 2002.

[16] Doug A. Bowman, Ernst Kruijff, and Joseph LaViola. *3D User Interfaces: Theory and Practice.* Addison Wesley, 1 edition, July 2004.

[17] John Brooke. *SUS: A "quick and dirty" usability scale.* In Patrick W. Jordan, B Thomas, Ian Lyall McClelland, and Bernard Weermeester, editors, Usability evaluation in industry, pages 189–194. Taylor & Francis, 1996.

[18] A.J. Bernheim Brush, Kerry Hammil, Steven Levi, Amy K. Karlson, James Scott, Raman Sarin, Andy Jacobs, Barry Bond, Oscar Murillo, Galen Hunt, and Mike Sinclair. *User experiences with activity-based navigation on mobile devices.* In Proceedings of the 12th international conference on Human computer interaction with mobile devices and services (MobileHCI 2010), pages 73—82, Lisbon, Portugal, 2010.

[19] Andreas Butz, Jörg Baus, Antonio Krüger, and Marco Lohse. *A hybrid indoor navigation system.* In Proceedings of the 6th international conference on Intelligent user interfaces, IUI '01, page 25–32, New York, NY, USA, 2001. ACM.

[20] Gartner S. Chapin. *Photo-Auto Maps.* Motor Car Supply Co., Chicago, IL, 1907.

[21] David M. Chen, Georges Baatz, Kevin Köser, Sam S. Tsai, Ramakrishna Vedantham, Timo Pylvänäinen, Kimmo Roimela, Xin Chen, Jeff Bach, Marc Pollefeys, Bernd Girod, and Radek Grzeszczuk. *City-scale landmark identification on mobile devices.* In Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, pages 737–744, Los Alamitos, CA, USA, 2011. IEEE Computer Society.

[22] Guanling Chen and David Kotz. *A Survey of Context-Aware Mobile Computing Research.* Technical report, Dartmouth College, Hanover, NH, USA, 2000.

[23] Keith Cheverst, Nigel Davies, Keith Mitchell, and Adrian Friday. *Experiences of developing and deploying a context-aware tourist guide: the GUIDE project.* In Proceedings of the 6th annual international conference on Mobile computing and networking, MobiCom '00, page 20–31, New York, NY, USA, 2000. ACM.

[24] Luca Chittaro and Stefano Burigat. *Augmenting audio messages with visual directions in mobile guides: an evaluation of three approaches.* In Proceedings of the 7th international conference on Human computer interaction with mobile devices and services, MobileHCI '05, page 107–114, New York, NY, USA, 2005. ACM.

[25] Luca Chittaro and Daniele Nadalutti. *Presenting evacuation instructions on mobile devices by means of location-aware 3D virtual environments.* In Proceedings of the 10th international conference on Human computer interaction with mobile devices and services, MobileHCI '08, page 395–398, New York, NY, USA, 2008. ACM.

[26] Luca Chittaro and Daniele Nadalutti. *A Mobile RFID-Based System for Supporting Evacuation of Buildings.* In Jobst Löffler and Markus Klann, editors, Mobile Response, page 22–31. Springer-Verlag, Berlin, Heidelberg, 2009.

[27] Jaewoo Chung and Chris Schmandt. *Going my way: a user-aware route planner.* In Proceedings of the 27th international conference on Human factors in computing systems, CHI '09, page 1899–1902, New York, NY, USA, 2009. ACM.

[28] Rudolph P. Darken and Barry Peterson. *Spatial Orientation, Wayfinding, and Representation.* Handbook of virtual environments: design, implementation and applications, pages 493–518, 2001.

[29] Andrew J. Davison, Walterio W. Mayol, and David W. Murray. *Real-Time Localisation and Mapping with Wearable Active Vision.* In Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR '03, page 18, Washington, DC, USA, 2003. IEEE Computer Society.

[30] Stephen DiVerdi, Jason Wither, and Tobias Höllerer. *Envisor: Online Environment Map Construction for Mixed Reality.* In Proceedings of the IEEE Virtual Reality Conference, pages 19–26. IEEE, 2008.

[31] Andreas Dünser, Mark Billinghurst, James Wen, Ville Lehtinen, and A. Nurminen. *Handheld AR for Outdoor Navigation.* In MobileHCI2011 - Workshop on Mobile Augmented Reality, 2011.

[32] Steven Feiner, Blair Macintyre, and Tobias Höllerer. *Wearing It Out: First Steps Toward Mobile Augmented Reality Systems.* In In First International Symposium on Mixed Reality (ISMR'99), page 363–377. Ohmsha (Tokyo)–Springer Verlag, 1999.

[33] Steven Feiner, Blair MacIntyre, Tobias Höllerer, and Anthony Webster. *A touring machine: prototyping 3D mobile augmented reality systems for exploring the urban environment.* In First International Symposium on Wearable Computers, pages 74–81. IEEE Comput. Soc, 1997.

[34] Martin A. Fischler and Robert C. Bolles. *Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography.* Commun. ACM, 24(6):381–395, June 1981.

[35] George W. Fitzmaurice. *Situated information spaces and spatially aware palmtop computers.* Commun. ACM, 36(7):39–49, July 1993.

[36] David C. Foyle, Anthony D. Andre, and Becky L. Hooey. *Situation Awareness in an Augmented Reality Cockpit: Design, Viewpoints and Cognitive Glue.* In Proceedings of Human-Computer Interaction International, pages 22–27, 2005.

[37] Peter Fröhlich, Matthias Baldauf, Marion Hagen, Stefan Suette, Dietmar Schabus, and Andrew L. Kun. *Investigating Safety Services on the Motorway: the Role of Realistic Visualization.* In 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Salzburg, Austria, 2011.

[38] Peter Fröhlich, Gerhard Obernberger, Rainer Simon, and Peter Reichl. *Exploring the design space of Smart Horizons.* In Proceedings of the 10th international conference on Human computer interaction with mobile devices and services, MobileHCI '08, page 363–366, New York, NY, USA, 2008. ACM.

[39] Peter Fröhlich, Rainer Simon, Lynne Baillie, and Hermann Anegg. *Comparing conceptual designs for mobile access to geo-spatial information.* In Proceedings of the 8th conference on Human-computer interaction with mobile devices and services, MobileHCI '06, page 109–112, New York, NY, USA, 2006. ACM.

[40] Steffen Gauglitz, Tobias Höllerer, and Matthew Turk. *Evaluation of Interest Point Detectors and Feature Descriptors for Visual Tracking.* International Journal of Computer Vision, 94(3):335–360, 2011.

[41] GeoVector. *GeoVector.* `http://www.geovector.com/`, 2012. [Online, last accessed on May 14, 2012].

[42] Andreas Girgensohn, Don Kimber, Jim Vaughan, Tao Yang, Frank Shipman, Thea Turner, Eleanor Rieffel, Lynn Wilcox, Francine Chen, and Tony Dunnigan. *DOTS: support for effective video surveillance.* In Proceedings of the 15th international conference on Multimedia, MULTIMEDIA '07, page 423–432, New York, NY, USA, 2007. ACM.

[43] Sarah E. Goldin and Perry W. Thorndyke. *Spatial Learning and Reasoning Skill.* Technical report, RAND Corporation, July 1981.

[44] Google. *Google Maps Navigation.* `http://www.google.com/mobile/navigation/`, 2012. [Online, last accessed on April 1, 2012].

[45] Raphael Grasset, Julian Looser, and Mark Billinghurst. *Transitional interface: concept, issues and framework.* In Proceedings of the 5th IEEE and ACM International

Symposium on Mixed and Augmented Reality, ISMAR '06, page 231–232, Washington, DC, USA, 2006. IEEE Computer Society. ACM ID: 1514241.

[46] Raphael Grasset, **Alessandro Mulloni**, Mark Billinghurst, and Dieter Schmalstieg. *Navigation Techniques in Augmented and Mixed Reality: Crossing the Virtuality Continuum.* In Borko Furht, editor, Handbook of Augmented Reality, pages 379–407. Springer New York, New York, NY, 2011.

[47] Sinem Guven and Steven Feiner. *Interaction Techniques for Exploring Historic Sites through Situated Media.* In IEEE Symposium on 3D User Interfaces (3DUI 2006), pages 111–118. IEEE, 2006.

[48] Sinem Guven and Steven Feiner. *Visualizing and navigating complex situated hypermedia in augmented and virtual reality.* In Proceedings of the IEEE/ACM International Symposium on Mixed and Augmented Reality, pages 155–158. IEEE, October 2006.

[49] Nate Hagbi, Oriel Bergig, Jihad El-Sana, Klara Kedem, and Mark Billinghurst. *Inplace Augmented Reality.* In 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, 2008. ISMAR 2008, pages 135–138. IEEE, September 2008.

[50] Drexel Hallaway, Steven Feiner, and Tobias HöLlerer. *Bridging the gaps: hybrid tracking for adaptive mobile augmented reality.* Applied Artificial Intelligence, 18(6):477–500, July 2004.

[51] Sandra G. Hart, Lowell E. Staveland, Peter A. Hancock, and Najmedin Meshkati. *Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research.* In Human Mental Workload, volume Volume 52, pages 139–183. North-Holland, 1988.

[52] Mary Hegarty. *Development of a self-report measure of environmental spatial ability.* Intelligence, 30(5):425–447, October 2002.

[53] Anders Henrysson, Mark Billinghurst, and Mark Ollila. *Face to Face Collaborative AR on Mobile Phones.* In Proceedings of the 4th IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR '05, page 80–89, Washington, DC, USA, 2005. IEEE Computer Society.

[54] Niels Henze, Enrico Rukzio, and Susanne Boll. *100,000,000 taps: analysis and improvement of touch performance in the large.* In Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, pages 133—142. ACM Press, 2011.

[55] Harlan Hile and Gaetano Borriello. *Information overlay for camera phones in indoor environments.* In Proceedings of the 3rd international conference on Location-and context-awareness, LoCA'07, page 68–84, Berlin, Heidelberg, 2007. Springer-Verlag.

[56] Harlan Hile, Radek Grzeszczuk, Alan Liu, Ramakrishna Vedantham, Jana Košecka, and Gaetano Borriello. *Landmark-Based Pedestrian Navigation with Enhanced Spatial Reasoning.* In Proceedings of the 7th International Conference on Pervasive Computing, Pervasive '09, page 59–76, Berlin, Heidelberg, 2009. Springer-Verlag.

[57] Harlan Hile, Alan Liu, Gaetano Borriello, Radek Grzeszczuk, Ramakrishna Vedantham, and Jana Košecka. *Visual Navigation for Mobile Devices.* IEEE Multimedia, 17(2):16–25, June 2010.

[58] Harlan Hile, Ramakrishna Vedantham, Gregory Cuellar, Alan Liu, Natasha Gelfand, Radek Grzeszczuk, and Gaetano Borriello. *Landmark-based pedestrian navigation from collections of geotagged photos.* In Proceedings of the 7th International Conference on Mobile and Ubiquitous Multimedia, MUM '08, page 145–152, New York, NY, USA, 2008. ACM.

[59] Stephen Hirtle, Kai-Florian Richter, Samvith Srinivas, and Robert Firth. *This is the tricky part: When directions become difficult.* Journal of Spatial Information Science, 1(1), July 2010.

[60] Simon Holland, David R. Morse, and Henrik Gedenryd. *AudioGPS: Spatial Audio Navigation with a Minimal Attention Interface.* Personal Ubiquitous Comput., 6(4):253–259, January 2002.

[61] Tobias Höllerer and Steven Feiner. *Mobile augmented reality.* Telegeoinformatics: Location-Based Computing and Services. Taylor and Francis Books Ltd., London, UK, 2004.

[62] Tobias Höllerer, Steven Feiner, and John Pavlik. *Situated documentaries: embedding multimedia presentations in the real world.* In The Third International Symposium on Wearable Computers (ISWC 1999), pages 79–86. IEEE Comput. Soc, 1999.

[63] Tobias Höllerer, Steven Feiner, Tachio Terauchi, Gus Rashid, and Drexel Hallaway. *Exploring MARS: developing indoor and outdoor user interfaces to a mobile augmented reality system.* Computers & Graphics, 23(6):779–785, December 1999.

[64] Tobias Höllerer, Drexel Hallaway, Navdeep Tinna, and Steven Feiner. *Steps Toward Accommodating Variable Position Tracking Accuracy in a Mobile Augmented Reality System.* In In Proceedings of AIMS 2001: Second Int. Workshop on Artificial Intelligence in Mobile Systems, volume 1, pages 31—37, 2001.

[65] Wolfgang Hürst and Tair Bilyalov. *Dynamic versus static peephole navigation of VR panoramas on handheld devices.* Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia, page 25:1–25:8, 2010.

[66] Wolfgang Hürst and Steffen Wittmer. *Navigating VR Panoramas on Mobile Devices.* In Information Visualisation, 2009 13th International Conference, pages 203–209. IEEE, July 2009.

[67] imaGinyze. *Augmented Driving.* `http://www.imaginyze.com/Site/Welcome.html`, 2012. [Online, last accessed on April 1, 2012].

[68] Hiroshi Ishii, Eran Ben-Joseph, John Underkoffler, Luke Yeung, Dan Chak, Zahra Kanji, and Ben Piper. *Augmented Urban Planning Workbench: Overlaying Drawings, Physical Models and Digital Simulation.* In Proceedings of the 1st International Symposium on Mixed and Augmented Reality, ISMAR '02, page 203, Washington, DC, USA, 2002. IEEE Computer Society.

[69] Bolan Jiang, Ulrich Neumann, and Suya You. *A robust hybrid tracking system for outdoor augmented reality.* In Virtual Reality, 2004. Proceedings. IEEE, page 3, March 2004.

[70] Antonio Ramón Jimenez, Fernando Seco, Carlos Prieto, and Jorge Guevara. *A comparison of Pedestrian Dead-Reckoning algorithms using a low-cost MEMS IMU.* In 2009 IEEE International Symposium on Intelligent Signal Processing, pages 37–42, Budapest, Hungary, August 2009.

[71] Matt Jones, Steve Jones, Gareth Bradley, Nigel Warren, David Bainbridge, and Geoff Holmes. *ONTRACK: Dynamically adapting music playback to support navigation.* Personal Ubiquitous Comput., 12(7):513–525, October 2008.

[72] Simon Julier, Yohan Baillot, Dennis Brown, and Marco Lanzagorta. *Information Filtering for Mobile Augmented Reality.* IEEE Comput. Graph. Appl., 22(5):12–15, September 2002.

[73] Simon Julier, Yohan Baillot, Marco Lanzagorta, Dennis Brown, and Lawrence Rosenblum. *BARS: Battlefield Augmented Reality System.* Technical report, Naval Research Lab, Washington DC, Advanced Information Technology, April 2001.

[74] Michael Kalkusch, Thomas Lidy, Michael Knapp, Gerhard Reitmayr, Hannes Kaufmann, and Dieter Schmalstieg. *Structured visual markers for indoor pathfinding.* In The First IEEE International Workshop on Augmented Reality Toolkit, page 8. IEEE, 2002.

[75] Hirokazu Kato and Mark Billinghurst. *Marker Tracking and HMD Calibration for a Video-Based Augmented Reality Conferencing System.* In Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality, IWAR '99, page 85, Washington, DC, USA, 1999. IEEE Computer Society.

[76] Michael Kenteris, Damianos Gavalas, and Daphne Economou. *Electronic mobile guides: a survey.* Personal and Ubiquitous Computing, 15(1):97–111, April 2010.

[77] Jongbae Kim and Heesung Jun. *Vision-based location positioning using augmented reality for indoor navigation.* IEEE Transactions on Consumer Electronics, 54(3):954–962, August 2008.

[78] SeungJun Kim and Anind K. Dey. *Simulated augmented reality windshield display as a cognitive mapping aid for elder driver navigation.* In Proceedings of the 27th international conference on Human factors in computing systems, CHI '09, page 133–142, New York, NY, USA, 2009. ACM.

[79] Georg Klein and Tom Drummond. *Robust Visual Tracking for Non-Instrumented Augmented Reality.* In Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR '03, page 113–, Washington, DC, USA, 2003. IEEE Computer Society.

[80] Georg Klein and David Murray. *Parallel Tracking and Mapping for Small AR Workspaces.* In Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR '07, page 1–10, Washington, DC, USA, 2007. IEEE Computer Society.

[81] Georg Klein and David Murray. *Parallel Tracking and Mapping on a camera phone.* In Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on, pages 83 –86, October 2009.

[82] Thomas Kolbe and Georg Gartner. *Augmented Videos and Panoramas for Pedestrian Navigation.* In Proceedings of the 2nd Symposium on Location Based Services and TeleCartography 2004 on 28.-29. January in Vienna, 2004.

[83] Johannes Kopf, Billy Chen, Richard Szeliski, and Michael Cohen. *Street Slide.* `http://research.microsoft.com/en-us/um/people/kopf/street_slide/`, 2010. [Online, last accessed on April 1, 2012].

[84] Johannes Kopf, Billy Chen, Richard Szeliski, and Michael Cohen. *Street slide: browsing street level imagery.* ACM Transactions on Graphics (TOG), 29:96:1–96:8, July 2010.

[85] Lynda J. Kramer. *Pathway design effects on synthetic vision head-up displays.* In Enhanced and Synthetic Vision, volume 5424, pages 61–72. SPIE, 2004.

[86] Christian Kray, Christian Elting, Katri Laakso, and Volker Coors. *Presenting route instructions on mobile devices.* In Proceedings of the 8th international conference on Intelligent user interfaces - IUI 2003, page 117, Miami, Florida, USA, 2003.

[87] Christian Kray and Gerd Kortuem. *Interactive Positioning Based on Object Visibility.* In Mobile Human-Computer Interaction - MobileHCI 2004, volume 3160, pages 276–287, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[88] Markus Kähäri and David J Murphy. *MARA – Sensor Based Augmented Reality System for Mobile Imaging Device.* Proceedings of the 5th Annual IEEE and ACM International Symposium on Mixed and Augmented Reality, pages 180–180, 2006.

[89] Richard B. Langley. *Dilution of precision.* GPS world, 10(5):52–59, 1999.

[90] Carol A. Lawton. *Gender differences in way-finding strategies: Relationship to spatial ability and spatial anxiety.* Sex Roles, 30(11-12):765–779, June 1994.

[91] Layar. *Layar.* `http://www.layar.com/`, 2012. [Online, last accessed on April 1, 2012].

[92] Juha Lehikoinen and Riku Suomela. *Accessing Context in Wearable Computers.* Personal Ubiquitous Comput., 6(1):64–74, January 2002.

[93] Zhang Lei and Paul Coulton. *A mobile geo-wand enabling gesture based POI search an user generated directional POI photography.* In Proceedings of the International Conference on Advances in Computer Enterntainment Technology, ACE '09, page 392–395, New York, NY, USA, 2009. ACM.

[94] Vincent Lepetit, Pascal Lagger, and Pascal Fua. *Randomized trees for real-time keypoint recognition.* In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, volume 2, pages 775–781, June 2005.

[95] Xiaowei Li, Changchang Wu, Christopher Zach, Svetlana Lazebnik, and Jan-Michael Frahm. *Modeling and Recognition of Landmark Image Collections Using Iconic Scene Graphs.* In Proceedings of the 10th European Conference on Computer Vision: Part I, ECCV '08, page 427–440, Berlin, Heidelberg, 2008. Springer-Verlag.

[96] Ting Liu, Andrew W. Moore, Er Gray, and Ke Yang. *An investigation of practical approximate nearest neighbor algorithms.* In Proceedings of Neural Information Processing Systems, page 825–832. MIT Press, 2004.

[97] Mark A. Livingston, Lawrence J. Rosenblum, Simon J. Julier, Dennis Brown, Yohan Baillot, J. Edward Swan, Joseph L. Gabbard, and Deborah Hix. *An augmented reality system for military operations in urban terrain.* Interservice / Industry Training, Simulation & Education Conference, page 89, 2002.

[98] Mark A. Livingston, Ed Swan, Joseph L. Gabbard, Tobias Höllerer, Deborah Hix, Simon J. Julier, Yohan Baillot, and Dennis Brown. *Resolving multiple occluded layers in augmented reality.* In The Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2003), pages 56–65. IEEE Comput. Soc, 2003.

[99] Mark A. Livingston, Ed Swan, Simon J. Julier, Yohan Baillot, Dennis Brown, Lawrence J. Rosenblum, Joseph L. Gabbard, Tobias Höllerer, and Deborah Hix. *Evaluating System Capabilities and User Performance in the Battlefield Augmented Reality System.* In Performance Metrics for Intelligent Systems Workshop, August 2004.

[100] David G. Lowe. *Distinctive Image Features from Scale-Invariant Keypoints.* Int. J. Comput. Vision, 60(2):91–110, November 2004.

[101] Kevin Lynch. *The Image of the City.* The MIT Press, June 1960.

[102] Markus Löchtefeld, Sven Gehring, Johannes Schöning, and Antonio Krüger. *PINwI: pedestrian indoor navigation without infrastructure.* In Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, NordiCHI '10, page 731–734, New York, NY, USA, 2010. ACM.

[103] Mary Madden, Lee Rainie, Pew Internet & American Life Project., and Pew Research Center. *Adults and cell phone distractions*, 2010.

[104] Sandy Martedi, Hideaki Uchiyama, Guillermo Enriquez, Hideo Saito, Tsutomu Miyashita, and Takenori Hara. *Foldable augmented maps.* In 2010 9th IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pages 312–312. IEEE, October 2010.

[105] Eladio Martin, Oriol Vinyals, Gerald Friedland, and Ruzena Bajcsy. *Precise indoor localization using smart phones.* In Proceedings of the international conference on Multimedia, MM '10, page 787–790, New York, NY, USA, 2010. ACM.

[106] Andrew J. May, Tracy Ross, Steven H. Bayer, and Mikko J. Tarkiainen. *Pedestrian navigation aids: information requirements and design implications.* Personal Ubiquitous Comput., 7(6):331–338, December 2003.

[107] Zeljko Medenica, Andrew L. Kun, Tim Paek, and Oskar Palinko. *Augmented reality vs. street views: a driving simulator study comparing two emerging navigation aids.* In Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, MobileHCI '11, page 265–274, New York, NY, USA, 2011. ACM.

[108] Liqiu Meng, Alexander Zipf, and Tumasch Reichenbacher. *Map-based mobile services: theories, methods and implementations.* Springer, 2005.

[109] Davide Merico and Roberto Bisiani. *Indoor Navigation with Minimal Infrastructure.* In 4th Workshop on Positioning, Navigation and Communication, 2007. WPNC '07, pages 141–144. IEEE, March 2007.

[110] metaio. *Junaio.* `http://www.junaio.com/`, 2012. [Online, last accessed on April 1, 2012].

[111] Pierre-Emmanuel Michon and Michel Denis. *When and Why Are Visual Landmarks Used in Giving Directions?* In Spatial Information Theory, volume 2205 of Lecture Notes in Computer Science, pages 292–305. Springer Berlin / Heidelberg, 2001.

[112] Microsoft. *Read/Write World.* http://readwriteworld.cloudapp.net/, 2012. [Online, last accessed on April 1, 2012].

[113] Alexandra Millonig and Katja Schechtner. *Developing Landmark-Based Pedestrian-Navigation Systems.* IEEE Transactions on Intelligent Transportation Systems, 8(1):43–49, March 2007.

[114] Mathias Mohring, Christian Lessig, and Oliver Bimber. *Video See-Through AR on Consumer Cell-Phones.* In Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR '04, page 252–253, Washington, DC, USA, 2004. IEEE Computer Society.

[115] Antoni Moore and Holger Regenbrecht. *The tangible augmented street map.* In Proceedings of the 2005 international conference on Augmented tele-existence, ICAT '05, page 249–250, New York, NY, USA, 2005. ACM.

[116] Ann Morrison, Giulio Jacucci, P. Pelto, Antti Juustila, and Gerhard Reitmayr. *Using locative games to evaluate hybrid technology.* British Computer Society, 2007.

[117] Ann Morrison, **Alessandro Mulloni**, Saija Lemmelä, Antti Oulasvirta, Giulio Jacucci, Peter Peltonen, Dieter Schmalstieg, and Holger Regenbrecht. *Collaborative use of mobile augmented reality with paper maps.* Computers & Graphics, 35(4):789–799, August 2011.

[118] Ann Morrison, Antti Oulasvirta, Peter Peltonen, Saija Lemmela, Giulio Jacucci, Gerhard Reitmayr, Jaana Näsänen, and Antti Juustila. *Like bees around the hive: a comparative study of a mobile augmented reality map.* In Proceedings of the 27th international conference on Human factors in computing systems - CHI '09, pages 1889—1898, Boston, MA, USA, 2009.

[119] **Alessandro Mulloni**, Andreas Dünser, and Dieter Schmalstieg. *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 161–170, New York, NY, USA, September 2010. ACM.

[120] **Alessandro Mulloni**, Hartmut Seichter, Andreas Dünser, Patrick Baudisch, and Dieter Schmalstieg. *360° Panoramic Overviews for Location-Based Services.* In Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems, CHI '12, page 2565–2568, New York, NY, USA, 2012. ACM.

[121] **Alessandro Mulloni**, Hartmut Seichter, and Dieter Schmalstieg. *Enhancing Handheld Navigation Systems with Augmented Reality.* In Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, Workshop on Mobile Augmented Reality, 2011.

[122] **Alessandro Mulloni**, Hartmut Seichter, and Dieter Schmalstieg. *Handheld augmented reality indoor navigation with activity-based instructions.* In Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, MobileHCI '11, page 211–220, New York, NY, USA, 2011. ACM.

[123] **Alessandro Mulloni**, Hartmut Seichter, and Dieter Schmalstieg. *User experiences with augmented reality aided navigation on phones.* In Proceedings of ISMAR 2011 (Poster), pages 229–230. IEEE, October 2011.

[124] **Alessandro Mulloni**, Hartmut Seichter, and Dieter Schmalstieg. *Indoor Navigation with Mixed-Reality World-in-Miniature Views.* In Proceedings of AVI 2012, 2012.

[125] **Alessandro Mulloni**, Daniel Wagner, Istvan Barakonyi, and Dieter Schmalstieg. *Indoor Positioning and Navigation with Camera Phones.* IEEE Pervasive Computing, 8(2):22–31, April 2009.

[126] Hans Jörg Müller, Johannes Schöning, and Antonio Krüger. *Mobile Map Interaction - Evaluation in an indoor scenario.* In GI Jahrestagung (2)'06, pages 403–410, 2006.

[127] Wolfgang Narzt, Gustav Pomberger, Alois Ferscha, Dieter Kolb, Reiner Müller, Jan Wieghardt, Horst Hörtner, and Christopher Lindinger. *Pervasive information acquisition for mobile AR-navigation systems.* In Fifth IEEE Workshop on Mobile Computing Systems & Applications, pages 13–20. IEEE, 2003.

[128] Richard A. Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J. Davison, Pushmeet Kohli, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. *KinectFusion: Real-time dense surface mapping and tracking.* In Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality, ISMAR '11, page 127–136, Washington, DC, USA, 2011. IEEE Computer Society.

[129] Richard L. Newman and Max Mulder. *Pathway displays: a literature review.* In Digital Avionics Systems Conference, 2003. DASC '03. The 22nd, volume 2, pages 91–10, October 2003.

[130] Nokia. *Ovi Maps Mobile.* `http://betalabs.nokia.com/ovi-maps`, 2012. [Online, last accessed on April 1, 2012].

[131] Moira Norrie and Beat Signer. *Overlaying Paper Maps with Digital Information Services for Tourists.* In Proceedings of the International Conference in Innsbruck, Austria, 2005, pages 23—33. Springer, 2005.

[132] Alex Olwal and Anders Henrysson. *LUMAR: A Hybrid Spatial Display System for 2D and 3D Handheld Augmented Reality.* In Proceedings of the 17th International Conference on Artificial Reality and Telexistence, pages 63–70. IEEE, November 2007.

[133] Antti Oulasvirta, Sara Estlander, and Antti Nurminen. *Embodied interaction with a 3D versus 2D mobile map.* Personal Ubiquitous Comput., 13(4):303–320, May 2009.

[134] Katharina Pentenrieder, Peter Meier, and Gudrun Klinker. *Analysis of Tracking Accuracy for Single-Camera Square-Marker-Based Tracking.* In Proc. Dritter Workshop Virtuelle und Erweiterte Realität der GI-Fachgruppe VR/AR, Koblenz, Germany, September 2006.

[135] Wayne Piekarski, Benjamin Avery, Bruce H. Thomas, and Pierre Malbezin. *Hybrid Indoor and Outdoor Tracking for Mobile 3D Mixed Reality.* In Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR '03, page 266–267, Washington, DC, USA, 2003. IEEE Computer Society.

[136] Wayne Piekarski, Benjamin Avery, Bruce H. Thomas, and Pierre Malbezin. *Integrated head and hand tracking for indoor and outdoor augmented reality.* In IEEE Virtual Reality, 2004. Proceedings, page 11. IEEE, March 2004.

[137] Wayne Piekarski, Bernard Gunther, and Bruce Thomas. *Integrating Virtual and Augmented Realities in an Outdoor Application.* In Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality, IWAR '99, page 45, Washington, DC, USA, 1999. IEEE Computer Society.

[138] Wayne Piekarski, David Hepworth, Victor Demczuk, Bruce Thomas, and Bernard Gunther. *A Mobile Augmented Reality User Interface for Terrestrial Navigation.* In 22nd Australasian Computer Science Conference, pages 122–133, 1999.

[139] Wayne Piekarski and Bruce H. Thomas. *Through-Walls Collaboration.* IEEE Pervasive Computing, 8(3):42–49, July 2009.

[140] Martin Pielot and Susanne Boll. *Tactile Wayfinder: Comparison of Tactile Waypoint Navigation with Commercial Pedestrian Navigation Systems.* In Patrik Floréen, Antonio Krüger, and Mirjana Spasojevic, editors, Pervasive Computing, volume 6030, pages 76–93. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[141] Martin Pielot, Niels Henze, and Susanne Boll. *Supporting map-based wayfinding with tactile cues.* In Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '09, page 23:1–23:10, New York, NY, USA, 2009. ACM.

[142] Martin Pielot, Benjamin Poppinga, Wilko Heuten, and Susanne Boll. *6th senses for everyone!: the value of multimodal feedback in handheld navigation aids.* In Proceedings of the 13th international conference on multimodal interfaces, ICMI '11, page 65–72, New York, NY, USA, 2011. ACM.

[143] Christian Pirchheim and Gerhard Reitmayr. *Homography-based planar mapping and tracking for mobile phones.* In Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality, ISMAR '11, page 27–36, Washington, DC, USA, 2011. IEEE Computer Society.

[144] Marina Plavšić, Markus Duschl, Marcus Tönnis, Heiner Bubb, and Gudrun Klinker. *Ergonomic Design and Evaluation of Augmented Reality Based Cautionary Warnings for Driving Assistance in Urban Environments.* In Proceedings of the International Ergonomics Association (IEA), August 2009. CD-ROM Proceedings.

[145] Benjamin Poppinga, Martin Pielot, and Susanne Boll. *Tacticycle: a tactile display for supporting tourists on a bicycle trip.* In Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '09, page 41:1–41:4, New York, NY, USA, 2009. ACM.

[146] Günther Pospischil, Martina Umlauft, and Elke Michlmayr. *Designing LoL@, a Mobile Tourist Guide for UMTS.* In Fabio Paternò, editor, Human Computer Inter-

action with Mobile Devices, volume 2411, pages 140–154. Springer Berlin Heidelberg, Berlin, Heidelberg, 2002.

[147] Arto Puikkonen, Ari-Heikki Sarjanoja, Merja Haveri, Jussi Huhtala, and Jonna Häkkilä. *Towards designing better maps for indoor navigation: experiences from a case study.* In Proceedings of the 8th International Conference on Mobile and Ubiquitous Multimedia, MUM '09, page 16:1–16:4, New York, NY, USA, 2009. ACM.

[148] Daniel Pustka, Manuel Huber, Christian Waechter, Florian Echtler, Peter Keitler, Joseph Newman, Dieter Schmalstieg, and Gudrun Klinker. *Automatic Configuration of Pervasive Sensor Networks for Augmented Reality.* IEEE Pervasive Computing, 10(3):68–79, July 2011.

[149] Qualcomm. *Vuforia.* `http://www.qualcomm.com/solutions/augmented-reality`, 2012. [Online, last accessed on April 1, 2012].

[150] Martin Raubal and Stephan Winter. *Enriching Wayfinding Instructions with Local Landmarks.* In Proceedings of the Second International Conference on Geographic Information Science, GIScience '02, page 243–259, London, UK, 2002. Springer-Verlag.

[151] Karl Rehrl, Elisabeth Häusler, and Sven Leitinger. *Comparing the Effectiveness of GPS-Enhanced Voice Guidance for Pedestrians with Metric- and Landmark-Based Instruction Sets.* In Lecture Notes in Computer Science, volume 6292, pages 189–203. Springer Berlin / Heidelberg, 2010.

[152] Karl Rehrl, Elisabeth Häusler, Renate Steinmann, Sven Leitinger, Daniel Bell, and Michael Weber. *Pedestrian Navigation with Augmented Reality, Voice and Digital Map: Results from a Field Study assessing Performance and User Experience.* In William Cartwright, Georg Gartner, Liqiu Meng, Michael P. Peterson, Georg Gartner, and Felix Ortag, editors, Advances in Location-Based Services, pages 3–20. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[153] Derek Reilly, Malcolm Rodgers, Ritchie Argue, Mike Nunes, and Kori Inkpen. *Marked-up maps: combining paper maps and electronic information resources.* Personal Ubiquitous Comput., 10(4):215–226, March 2006.

[154] Gerhard Reitmayr and Tom Drummond. *Going out: robust model-based tracking for outdoor augmented reality.* In Mixed and Augmented Reality, 2006. ISMAR 2006. IEEE/ACM International Symposium on, pages 109—118, October 2006.

[155] Gerhard Reitmayr and Tom Drummond. *Initialisation for Visual Tracking in Urban Environments.* In Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on, pages 161—172, November 2007.

[156] Gerhard Reitmayr, Ethan Eade, and Tom Drummond. *Localisation and interaction for augmented maps.* In Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality, 2005. Proceedings, pages 120—129. IEEE, October 2005.

[157] Gerhard Reitmayr and Dieter Schmalstieg. *Location based applications for mobile augmented reality.* In Proceedings of the Fourth Australasian user interface conference on User interfaces 2003 - Volume 18, AUIC '03, page 65–73, Darlinghurst, Australia, Australia, 2003. Australian Computer Society, Inc.

[158] Gerhard Reitmayr and Dieter Schmalstieg. *Collaborative augmented reality for outdoor navigation and information browsing.* In Proceedings of the Second Symposium on Location Based Services and TeleCartography, volume 66, page 31–41, 2004.

[159] Miguel Ribo, Peter Lang, Harald Ganster, Markus Brandner, Christoph Stock, and Axel Pinz. *Hybrid tracking for outdoor augmented reality applications.* Computer Graphics and Applications, IEEE, 22(6):54—63, December 2002.

[160] George G. Robertson, Jock D. Mackinlay, and Stuart K. Card. *Cone Trees: animated 3D visualizations of hierarchical information.* In Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology, CHI '91, page 189–194, New York, NY, USA, 1991. ACM.

[161] Simon Robinson, Matt Jones, Parisa Eslambolchilar, Roderick Murray-Smith, and Mads Lindborg. *"I did it my way": moving away from the tyranny of turn-by-turn pedestrian navigation.* In Proceedings of the 12th international conference on Human computer interaction with mobile devices and services, MobileHCI '10, page 341–344, New York, NY, USA, 2010. ACM.

[162] Michael Rohs and Beat Gfeller. *Using Camera-Equipped Mobile Phones for Interacting with Real-World Objects.* In Advances in Pervasive Computing, page 265–271, 2004.

[163] Michael Rohs and Antti Oulasvirta. *Target acquisition with camera phones when used as magic lenses.* In Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, CHI '08, page 1409–1418, New York, NY, USA, 2008. ACM.

[164] Michael Rohs, Antti Oulasvirta, and Tiia Suomalainen. *Interaction with magic lenses: real-world validation of a Fitts' Law model.* In Proceedings of the 2011 annual conference on Human factors in computing systems, CHI '11, page 2725–2728, New York, NY, USA, 2011. ACM.

[165] Michael Rohs, Robert Schleicher, Johannes Schöning, Georg Essl, Anja Naumann, and Antonio Krüger. *Impact of item density on the utility of visual context in magic lens interactions.* Personal Ubiquitous Comput., 13(8):633–646, November 2009.

[166] Michael Rohs, Johannes Schöning, Martin Raubal, Georg Essl, and Antonio Krüger. *Map navigation with mobile devices: virtual versus physical movement with and without visual context.* In Proceedings of the 9th international conference on Multimodal interfaces, ICMI '07, page 146–153, New York, NY, USA, 2007. ACM.

[167] Edward Rosten and Tom Drummond. *Machine Learning for High-Speed Corner Detection.* In ECCV 2006, volume 3951, pages 430–443, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[168] Sonja Rümelin, Enrico Rukzio, and Robert Hardy. *NaviRadar: a novel tactile information display for pedestrian navigation.* In Proceedings of the 24th annual ACM symposium on User interface software and technology, UIST '11, page 293–302, New York, NY, USA, 2011. ACM.

[169] Christian Sandor, Andrew Cunningham, Ulrich Eck, Donald Urquhart, Graeme Jarvis, Arindam Dey, Sebastien Barbier, Michael R. Marner, and Sang Rhee. *Egocentric space-distorting visualizations for rapid environment exploration in mobile mixed reality.* In Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality, ISMAR '09, page 211–212, Washington, DC, USA, 2009. IEEE Computer Society.

[170] Christian Sandor, Arindam Dey, Andrew Cunningham, Sebastien Barbier, Ulrich Eck, Donald Urquhart, Michael R. Marner, Graeme Jarvis, and Rhee Sang. *Egocentric space-distorting visualizations for rapid environment exploration in mobile*

mixed reality. In 2010 IEEE Virtual Reality Conference (VR), pages 47–50. IEEE, March 2010.

[171] Kiyohide Satoh, Mahoro Anabuki, Hiroyuki Yamamoto, and Hideyuki Tamura. *A hybrid registration method for outdoor augmented reality.* In Augmented Reality, 2001. Proceedings. IEEE and ACM International Symposium on, pages 67—76, 2001.

[172] Gerhard Schall, Erick Mendez, Ernst Kruijff, Eduardo Veas, Sebastian Junghanns, Bernhard Reitinger, and Dieter Schmalstieg. *Handheld Augmented Reality for underground infrastructure visualization.* Personal Ubiquitous Comput., 13(4):281–291, May 2009.

[173] Gerhard Schall, **Alessandro Mulloni**, and Gerhard Reitmayr. *North-centred orientation tracking on mobile phones.* In ISMAR 2010, pages 267–268, Seoul, Korea (South), October 2010.

[174] Torben Schinke, Niels Henze, and Susanne Boll. *Visualization of off-screen objects in mobile augmented reality.* In Proceedings of the 12th international conference on Human computer interaction with mobile devices and services, MobileHCI '10, page 313–316, New York, NY, USA, 2010. ACM.

[175] Björn Schwerdtfeger and Gudrun Klinker. *Supporting order picking with Augmented Reality.* In Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR '08, page 91–94, Washington, DC, USA, 2008. IEEE Computer Society.

[176] Björn Schwerdtfeger, Rupert Reif, Willibald A. Günthner, Gudrun Klinker, Daniel Hamacher, Lutz Schega, Irina Böckelmann, Fabian Doil, and Johannes Tümler. *Pick-by-Vision: A first stress test.* In 8th IEEE International Symposium on Mixed and Augmented Reality, 2009. ISMAR 2009, pages 115–124. IEEE, October 2009.

[177] Johannes Schöning, Antonio Krüger, Keith Cheverst, Michael Rohs, Markus Löchtefeld, and Faisal Taher. *PhotoMap: using spontaneously taken images of public maps for pedestrian navigation tasks on mobile devices.* In Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '09, page 14:1–14:10, New York, NY, USA, 2009. ACM.

[178] Johannes Schöning, Antonio Krüger, and Hans Jörg Müller. *Interaction of mobile camera devices with physical maps.* In Adjunct Proceeding of the Fourth International Conference on Pervasive Computing, page 121–124, 2006.

[179] Johannes Schöning, Michael Rohs, Sven Kratz, Markus Löchtefeld, and Antonio Krüger. *Map torchlight: a mobile augmented reality camera projector unit.* In Proceedings of the 27th international conference extended abstracts on Human factors in computing systems, CHI EA '09, page 3841–3846, New York, NY, USA, 2009. ACM.

[180] Alexander W. Siegel and Sheldon H. White. *The development of spatial representations of large-scale environments.* Advances in Child Development and Behavior, 10:9–55, 1975. PMID: 1101663.

[181] Rainer Simon, Peter Fröhlich, and Thomas Grechenig. *GeoPointing: evaluating the performance of orientation-aware location-based interaction under real-world conditions.* Journal of Location Based Services, 2(1):24–40, 2008.

[182] Iryna Skrypnyk and David G. Lowe. *Scene Modelling, Recognition and Tracking with Invariant Image Features.* In Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR '04, page 110–119, Washington, DC, USA, 2004. IEEE Computer Society.

[183] Randall C. Smith and Peter Cheeseman. *On the representation and estimation of spatial uncertainly.* Int. J. Rob. Res., 5(4):56–68, December 1986.

[184] Making Virtual Solid. *Virtual Cable.* `http://www.mvs.net/`, 2012. [Online, last accessed on April 1, 2012].

[185] Aaron Stafford, Bruce H. Thomas, and Wayne Piekarski. *Efficiency of techniques for mixed-space collaborative navigation.* In Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR '08, page 181–182, Washington, DC, USA, 2008. IEEE Computer Society.

[186] Aaron Stafford, Bruce H. Thomas, and Wayne Piekarski. *Comparison of techniques for mixed-space collaborative navigation.* In Proceedings of the Tenth Australasian Conference on User Interfaces - Volume 93, AUIC '09, page 61–70, Darlinghurst, Australia, Australia, 2009. Australian Computer Society, Inc.

[187] Richard Stoakley, Matthew J Conway, and Randy Pausch. *Virtual reality on a WIM: interactive worlds in miniature.* In Proceedings of the SIGCHI conference on Human factors in computing systems, CHI '95, page 265–272, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co. ACM ID: 223938.

[188] Ryuhei Tenmoku, Masayuki Kanbara, and Naokazu Yokoya. *A wearable augmented reality system using positioning infrastructures and a pedometer.* In Seventh IEEE International Symposium on Wearable Computers, 2003. Proceedings, pages 110–117. IEEE, October 2005.

[189] Bruce Thomas, Wayne Piekarski, David Hepworth, Bernard Gunther, and Victor Demczuk. *A Wearable Computer System with Augmented Reality to Support Terrestrial Navigation.* In Proceedings of the 2nd IEEE International Symposium on Wearable Computers, ISWC '98, page 168, Washington, DC, USA, 1998. IEEE Computer Society.

[190] Perry W. Thorndyke and Barbara Hayes-Roth. *Differences in spatial knowledge acquired from maps and navigation.* Cognitive Psychology, 14(4):560–589, October 1982.

[191] Yoshitaka Tokusho and Steven Feiner. *Prototyping an Outdoor Mobile Augmented Reality Street View Application.* ISMAR 2009, 2(c):3–5, 2009.

[192] TomTom. *TomTom for iOS.* `http://www.tomtom.com/en_gb/products/mobile-navigation/`, 2012. [Online, last accessed on April 1, 2012].

[193] Takahiro Tsuda, Haruyoshi Yamamoto, Yoshinari Kameda, and Yuichi Ohta. *Visualization methods for outdoor see-through vision.* In Proceedings of the 2005 international conference on Augmented tele-existence, ICAT '05, page 62–69, New York, NY, USA, 2005. ACM.

[194] Marcus Tönnis and Gudrun Klinker. *Effective control of a car driver's attention for visual and acoustic guidance towards the direction of imminent dangers.* In Proceedings of the 5th IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR '06, page 13–22, Washington, DC, USA, 2006. IEEE Computer Society.

[195] Marcus Tönnis and Gudrun Klinker. *Augmented 3D Arrows Reach Their Limits In Automotive Environments.* In Xiangyu Wang and Marc Aurel Schnabel, editors,

Mixed Reality In Architecture, Design And Construction, pages 185–202. Springer Netherlands, Dordrecht, 2009.

[196] Marcus Tönnis, Christian Lange, and Gudrun Klinker. *Visual Longitudinal and Lateral Driving Assistance in the Head-Up Display of Cars.* In Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, pages 1–4. IEEE, November 2007.

[197] Jan B. F. Van Erp, Hendrik A. H. C. Van Veen, Chris Jansen, and Trevor Dobbins. *Waypoint navigation with a vibrotactile waist belt.* ACM Trans. Appl. Percept., 2(2):106–117, April 2005.

[198] Eduardo Veas, **Alessandro Mulloni**, Ernst Kruijff, Holger Regenbrecht, and Dieter Schmalstieg. *Techniques for view transition in multi-camera outdoor environments.* In Proceedings of Graphics Interface 2010, GI '10, page 193–200, Toronto, Ont., Canada, Canada, 2010. Canadian Information Processing Society.

[199] Daniel Wagner, Tobias Langlotz, and Dieter Schmalstieg. *Robust and unobtrusive marker tracking on mobile phones.* In 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, 2008. ISMAR 2008., pages 121—124, September 2008.

[200] Daniel Wagner, **Alessandro Mulloni**, Tobias Langlotz, and Dieter Schmalstieg. *Real-time panoramic mapping and tracking on mobile phones.* In 2010 IEEE Virtual Reality Conference (VR), pages 211–218, Boston, MA, USA, March 2010.

[201] Daniel Wagner, Gerhard Reitmayr, **Alessandro Mulloni**, Tom Drummond, and Dieter Schmalstieg. *Pose tracking from natural features on mobile phones.* In 2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, pages 125–134, Cambridge, UK, September 2008.

[202] Daniel Wagner, Gerhard Reitmayr, **Alessandro Mulloni**, Tom Drummond, and Dieter Schmalstieg. *Real-Time Detection and Tracking for Augmented Reality on Mobile Phones.* IEEE Transactions on Visualization and Computer Graphics, 16(3):355–368, May 2010.

[203] Daniel Wagner and Dieter Schmalstieg. *First steps towards handheld augmented reality.* In Seventh IEEE International Symposium on Wearable Computers, 2003. Proceedings, pages 127—135. IEEE, October 2005.

[204] Benjamin Walther-Franks and Rainer Malaka. *Evaluation of an Augmented Photograph-Based Pedestrian Navigation System.* In Proceedings of the 9th international symposium on Smart Graphics, SG '08, page 94–105, Berlin, Heidelberg, 2008. Springer-Verlag.

[205] Rainer Wasinger and Antonio Krüger. *Multi-modal Interaction with Mobile Navigation Systems.* Special Journal Issue Conversational User Interfaces, it - Information Technology, 46(6):322–331, 2004.

[206] Christopher D. Wickens, Amy L. Alexander, William J. Horrey, Ashley Nunes, and Thomas J. Hardy. *Traffic and Flight Guidance Depiction on a Synthetic Vision System Display: The Effects of Clutter on Performance and Visual Attention Allocation.* Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 48(1):218–222, September 2004.

[207] Wikitude. *Wikitude.* `http://www.wikitude.com/`, 2012. [Online, last accessed on April 1, 2012].

[208] Wikitude. *Wikitude Drive.* `http://www.wikitude.com/tour/wikitude-drive`, 2012. [Online, last accessed on April 1, 2012].

[209] Jason Wither and Tobias Höllerer. *Pictorial Depth Cues for Outdoor Augmented Reality.* In Proceedings of the Ninth IEEE International Symposium on Wearable Computers, ISWC '05, page 92–99, Washington, DC, USA, 2005. IEEE Computer Society.

[210] Jason Wither, Yun-Ta Tsai, and Ronald Azuma. *Indirect augmented reality.* Computers & Graphics, 35(4):810–822, August 2011.

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

[212] Paul A. Zandbergen and Sean J. Barbeau. *Positional Accuracy of Assisted GPS Data from High-Sensitivity GPS-Enabled Mobile Phones.* The Journal of Navigation, 64(03):381–399, 2011.

[213] Ana Zanella, Marianne Sheelagh Therese Carpendale, and Michael Rounding. *On the effects of viewing cues in comprehending distortions.* In Proceedings of the second

Nordic conference on Human-computer interaction, NordiCHI ’02, page 119–128, New York, NY, USA, 2002. ACM.

[214] Jiang Yu Zheng. *Digital route panoramas.* IEEE Multimedia, 10(3):57– 67, September 2003.

[215] Ferdinand R. H. Zijlstra. *Efficiency in work behaviour: a design approach for modern tools.* Technical report, Roe, R.A., prof.dr. (promotor), Mulder, G., prof.dr. (promotor), 1993.

[216] Stefanie Zollmann, Denis Kalkofen, Erick Mendez, and Gerhard Reitmayr. *Image-based ghostings for single layer occlusions in augmented reality.* In 2010 9th IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pages 19–26. IEEE, October 2010.

[217] Mustafa Özuysal, Pascal Fua, and Vincent Lepetit. *Fast Keypoint Recognition in Ten Lines of Code.* In Computer Vision and Pattern Recognition, 2007. CVPR ’07. IEEE Conference on, pages 1—8, June 2007.