CARL VON OSSIETZKY universität OLDENBURG

Fakultät II – Informatik, Wirtschafts- und Rechtswissenschaften
Department für Informatik

Camera-based Mobile Interaction with Physical Objects

Dissertation zur Erlangung des Grades eines
Doktors der Naturwissenschaften

vorgelegt von

Dipl.-Inform. Niels Henze

Gutachter:

Prof. Dr. Susanne Boll, Universität Oldenburg
Prof. Dr. Enrico Rukzio, Universität Ulm

Tag der Disputation: 19. Juli 2012

Abstract

Despite the rise of the smartphone and the pervasiveness of digital services, interacting with physical objects is fortunately still essential in our daily life. Printed books are more convenient than E-books, points of interests become particularly interesting if one stands in front of them, and sales figures for music CDs shows that four times more physical albums are sold than digital ones. Digital services, however, enable use cases that physical media fails to offer. While the digital world is increasingly getting dynamic and interactive – physical media remains static. One cannot read the latest gossip about Madonna with just a CD in the hand. Instead one must use a search engine to find Madonna's MySpace page for a virtual meet-up with other fans. There is no easily accessible link that closes the gap between physical objects and digital services. Camera-based mobile interaction techniques enable to select physical objects to access related digital services. In the simplest case the user takes a photo with a mobile phone and receives relevant information about a photographed object. This dissertation presents the first comprehensive investigation of mobile camera-based interaction techniques with a focus on the users' perspective. Addressed research questions are: Do these interaction techniques provide a benefit for the user? Which interaction technique should be preferred? How to design the user interface for handheld AR systems? How should physical objects beyond the phone's screen be visualized? Answers to these questions can guide interaction designers when developing mobile interaction with the real world applications and inform the investigation of appropriate algorithms for such systems.

This thesis defines the research field described by the notion 'Mobile Interaction with the Real World' and provide a categorization of previous work. Three distinct types of camera-based interaction techniques are identified and analysed through user studies. We design, implement, and evaluate the interaction techniques Point & Shoot, Continuous Pointing, and handheld Augmented Reality. We provide evidence that camera-based interaction techniques can be more efficient and are preferred by users compared to manual techniques. We further show that camera-based interaction is not only usable by tech-savvy early adaptors but also by average consumers. Comparing the three interaction techniques we provide evidence that handheld Augmented Reality is preferred and reduces the perceived task load. While handheld Augmented Reality is widely studied with a technical focus this work is the first that investigates the interface design of handheld Augmented Reality applications. Using participatory design we develop design alternatives to augment printed photobooks and music CDs. We show that an object-aligned augmentation is more efficient for information presentation and, in contrast, also show that input controls should be aligned to the device's display. We further investigate the visualization of off-screen objects for handheld Augmented Reality, which is of particular importance considering the small screen size of mobile devices. We conduct a controlled experiment by publishing an application to an application store to revise previously proposed visualizations. We show that an off-screen visualization has a high impact on users' performance when interacting with augmented maps and provide evidence that an off-screen visualization is more usable than traditional techniques.

Zusammenfassung

Trotz der großen Verbreitung von Smartphones und digitalen Diensten ist die Interaktion mit physischen Objekten glücklicherweise immer noch von essenzieller Bedeutung in unserem Alltag. Gedruckte Bücher sind bequemer zu lesen als E-Books, Sehenswürdigkeiten bekommen eine besondere Bedeutung wenn wir vor ihnen stehen und es werden mehr physische Musik CDs als digitale Alben verkauft. Digitale Dienste eröffnen jedoch Möglichkeiten, die physische Medien nicht bieten. Während die digitale Welt immer dynamischer und interaktiver wird, bleiben physische Medien statisch. Nur mit einer CD ist es nicht möglich, die letzten Neuigkeiten über Madonna lesen. Stattdessen muss eine Suchmaschine verwenden werden um Madonnas MySpace Seite für ein virtuelles Treffen mit anderen Fans zu finden. Es gibt keine leicht zugängliche Verbindung, welche die Lücke zwischen physischen Objekten und digitalen Diensten schließt. Kamerabasierte mobile Interaktionstechniken ermöglichen es, physische Objekte auszuwählen, um auf digitale Dienste zuzugreifen. Im einfachsten Fall erstellt der Benutzer ein Foto mit einem Mobiltelefon und erhält relevante Informationen über das fotografierte Objekt. Diese Dissertation präsentiert die erste umfassende Untersuchung von mobilen kamerabasierten Interaktionstechniken unter besonderer Berücksichtigung der Nutzerperspektive. Adressierte Forschungsfragen sind: Bieten diese Interaktionstechniken einen Vorteil für den Benutzer? Welche Interaktionstechnik sollte bevorzugt werden? Wie sollte die Benutzeroberfläche für handheld Augmented Reality Systeme gestaltet sein? Wie sollten physische Objekte außerhalb des Bildschirms visualisiert werden? Antworten auf diese Fragen können Interaktionsdesigner bei der Entwicklung von entsprechenden Anwendungen anleiten und sie ermöglichen die Untersuchung von geeigneten Algorithmen für solche Systeme.

Diese Arbeit definiert den durch den Begriff "Mobile Interaction with the Real World" beschriebenen Forschungsbereich und bietet eine Kategorisierung bisheriger Arbeiten. Drei verschiedene kamerabasierte Interaktionstechniken werden identifiziert und durch Studien analysiert. Wir entwerfen, implementieren und evaluieren die Interaktionstechniken Point & Shoot, Continuous Pointing und handheld Augmented Reality. Es wird gezeigt, dass kamerabasierte Interaktionstechniken effizienter sein können und von Nutzern im Vergleich zu manuellen Techniken bevorzugt werden. Weiter zeigen wir, dass kamerabasierte Interaktion nicht nur von technisch versierten Benutzern, sondern auch vom durchschnittlichen Verbraucher verwendbar ist. Durch einen Vergleich der drei Interaktionstechniken wird belegt, dass handheld Augmented Reality bevorzugt wird und die gefühlte Arbeitsbelastung senken kann. Während handheld Augmented Reality bereits umfassend mit einem technischen Schwerpunkt untersucht wurde, ist diese Arbeit die erste, die das Interface-Design von handheld Augmented Reality Anwendungen betrachtet. Mittels partizipativen Designs werden Designalternativen für die Augmentierung von gedruckten Fotobüchern und Musik CDs entwickelt. Wir zeigen, dass eine an den physischen Objekten ausgerichtete Augmentation effizienter für die Darstellung von Informationen ist. Im Gegensatz dazu, wird auch gezeigt, dass Steuerelemente am Bildschirm des mobilen Geräts ausgerichtet sein sollten. Da die Visualisierung von

Objekten außerhalb des Fokus des Benutzers, insbesondere aufgrund der geringen Bildschirmgröße von mobilen Geräten, von besonderer Bedeutung ist, wird diese untersucht. Durch Veröffentlichung von mobilen Anwendungen werden erstmals kontrollierte Experimente zum Vergleich existierender Visualisierungstechniken mit einer großen Anzahl an Teilnehmern durchgeführt. Es wird außerdem gezeigt, dass eine Off-Screen Visualisierung eine große Auswirkung auf die Verwendbarkeit von augmentierten Karten hat und es wird nachgewiesen, dass die entwickelte Visualisierungstechnik eine höhere Benutzbarkeit als etablierte Techniken hat.

Acknowledgements

I thank my supervisor Susanne Boll for her guidance and the critical discussions but most importantly I want to thank her for the freedom she gave me. Pursuing my PhD in her group shaped the way I think and I am grateful for everything I learned from working with her. I thank Palle Klante, Wilko Heuten, and Enrico Rukzio for sharing their perspective on research with me and their encouragement. Special thanks to Enrico Rukzio for agreeing to referee my thesis and keeping me motivated. I thank my colleagues and friends from OFFIS, from the University of Oldenburg, and from the projects I had the chance to work for. Most notably I want to thank Martin Pielot, Benjamin Poppinga, Tobias Hesselmann, Jutta Fortmann, and Dirk Ahlers for being valued co-authors as well as good friends. I hope they learned from me as much as I learned from them. I also thank the students I had the chance to work with. In particular, I want to thank Torben Schinke and Andreas Löcken for their excellent work and especially for questioning my assumptions. I'm grateful to be able to work with all the great colleagues from the different projects I worked for, in particular, from the EU projects InterMedia and Enabled. I thank Albrecht Schmidt for the opportunity to finalize my thesis in his outstanding group. I am grateful that Alireza Sahami and my other new colleagues welcomed me and thank them for widening my perspective. Last but not least I thank all my parents and my sisters for making me the person I am. Finally, I am in debt to Jenni for her understanding and patience. Thank you for loving and supporting me more than I deserve.

Contents

1 Introduction

At least since the rise of the internet a truly vast amount of digital content became publicly available. Services such as Electronic Yellow Pages make information about local businesses globally available. Online stores like Amazon.com offer millions of products and provide extensive descriptions about them. United Kingdom's data.gov.uk and similar initiatives aim at open up all non-personal data acquired for official purposes [ST10]. Probably most impressive is the amount of user-generated content that emerged in recent years. Wikipedia states that 18 million articles in 279 languages about a broad range of subjects have been created [The11], users have uploaded billions of photos to Facebook [Bea08], and the Internet Movie Database (IMDb) contains detailed descriptions of 1.75 million movies [Nee10]. Today, the information provided by the World Wide Web is almost inconceivable in our contemporary society.

Large quantity of the available content refers to physical entities or geographic locations. Analysing what people search on the Web using common search engines shows that at least 13% of all search terms target geographic entities [GAMS08] and 20% of the queries are about people, places, and things [SJWS02]. For mobile users even over 30% of all queries are of geographical nature [CS09]. Physical entities are triggers for an information need. Standing in front of a film poster, for example, one might want to check film reviews before deciding to go to the cinema. We could watch the movies' trailers, read the directors' Wikipedia articles, and view the show times to find out if a movie meets our demands. Similarly, tourists want to access information about the sights they visit, bargain hunters are interested in product descriptions and user generated reviews even while shopping in physical stores, and reading paper documents one might want to examine background information.

Making information ubiquitary available in the form of digital content is a driving force behind the recent smartphone boom. Retrieving information about a movie while standing in front of its advertisement poster is possible today. Analysis of mobile information needs shows, however, that mobile needs differ significantly from general Web needs [CS09]. Retrieving information about a physical entity and satisfying the information need triggered by people, places, and things can be challenging especially while on the move. Physical entities do not provide a link to click. Despite the pervasiveness of mobile phones there is no easily accessible connection between physical objects and related digital services.

An approach to narrow this gap is augmenting physical objects with the interactivity, personalization, and real-time features provided by digital content using handheld devices. Mobile phones and other mobile devices are made aware of the physical objects in the user's proximity. Researchers developed according prototypes that enable to simply touch a printed book with a phone to access the book's website [WFGH99], pointing with a phone at sights and take a photo to retrieve further information [DCDH05a], or visually augmenting physical objects [Fit93, RN95] through a phone's screen. Different technical solutions that enable mobile devices to sense nearby physical objects have been proposed and developed. Most approaches either equip physical objects with electronic tags such as NFC or analyse images from mobile phones' cameras to recognize objects. Equipping a reasonable number of objects with electronic tags like RFID, NFC or Bluetooth emitters, however, requires major investment into the infrastructure. Camera-based approaches, which solely rely on the input from a camera that is already integrated in almost all smartphones available today, are therefore especially promising. The camera image is used to recognize physical objects that are present in front of the phone using visual markers or vision-based object recognition. The main advantage of vision-based object recognition is that in contrast to other approaches it is not required to alter the physical objects in any way. A camera-based approach is therefore the only technical solution that is feasible for a wide range of objects – even today.

In recent years a number of commercial mobile applications emerged that enable to access information and digital services by using physical objects as an anchor to the service. SnapTell, Nokia Point & Find, and Google Goggles are some examples of mobile applications that let users point their smartphone's camera at objects they want to know more about. By taking a photo of a physical object that is transmitted to a remote server the user receives related information and links to digital services. Using SnapTell, for example, the user can take a photo of a CD to receive links to the band's Wikipedia article, YouTube videos, and online stores. Another example are handheld Augmented Reality (handheld AR) applications, such as Layar, the Wikitude World Browser, and ZipRealty, that estimate a phone's position and orientation using GPS and the phone's compass to augmented the camera image with localised information. Such applications have been installed several million times and these sheer numbers show not only the vital interest of consumers but also the commercial potential.

Current commercial applications using a camera-based approach fulfil the functional requirement to provide content related to physical objects. The interaction design and the design of the user interface is, however, crucial when developing interactive systems. This is particularly true for mobile applications that must take the "fragmented nature of attentional resources in mobile HCI" [OTRK05] into account and, of course, also applies to camera-based mobile applications. It remains unclear if camera-based interaction techniques can be as efficient and effective as established mobile interaction techniques. In addition, different camera-based techniques can be used but it has not been studied which should be preferred when developing mobile applications. How to design the user interface on top of a particular interaction technique has almost been neglect by the research community and it has not been investigated how the mobile device should indicate the availability of "interactive" physical objects in the user's surrounding.

In the remainder of this chapter, we will motivate the research presented in this thesis in the field of mobile human-computer interaction in Section 1.1. Section 1.2 presents the addressed research challenges and the contribution of our work. Section 1.3 provides an outline of this thesis and an overview about the publications that contribute to it.

1.1 Scenario and Use Cases

In our daily life we are surrounded by physical objects that can serve as a link to digital content. One example are film posters that advertise movies playing in the local cinema. A person that spots a poster, such as the one sketched in Figure 1.1, can usually get basic information about the movie from the poster. The poster in Figure 1.1, for example, shows the name of the movie and announces that it is coming soon to the cinemas. If the poster attracts our attention we might want to learn more about the movie but the poster alone cannot provide additional information. With one of the recent visual search applications installed on our smartphone we could use the physical poster as an anchor to digital content that provides further information. A user could, for instance, create a photo of the poster using the visual search application Google Goggles. The mobile application would upload the photo to a server where it is analysed and hopefully recognized. If the poster is recognized the server sends metadata back to the mobile client. In the case of Google Goggles the user has to select the most appropriate result if multiple results are determined (see Figure 1.2). Google Goggles, as shown in Figure 1.2, provides some basic information about the movie. In addition, the option to initiate a product search as well as links to Wikipedia, the IMDb, and websites with movie reviews are available.

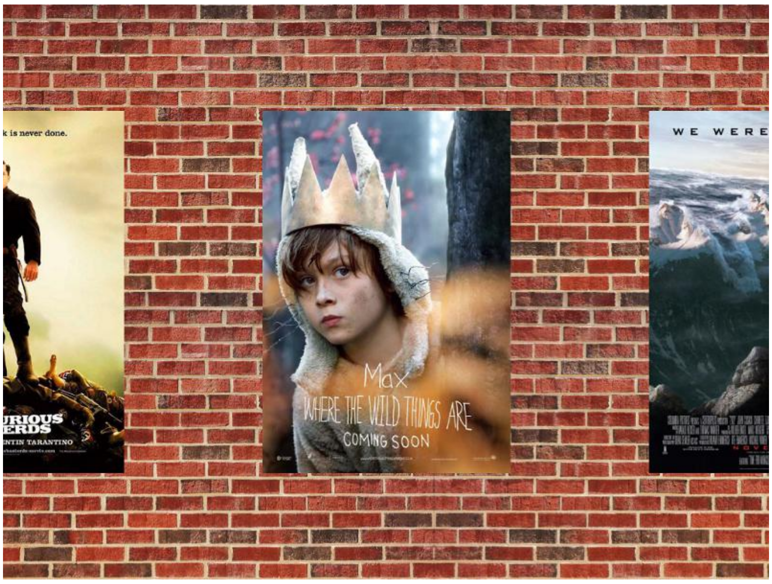
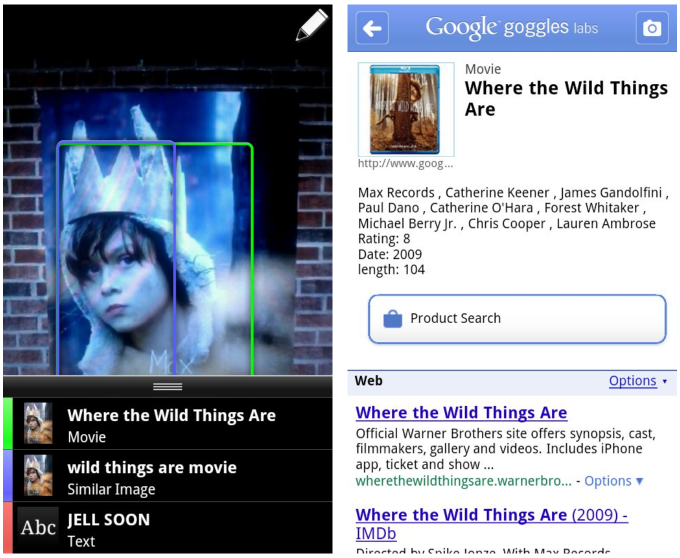


Figure 1.1: Sketch of three film posters that could be connected to related digital content such as a description, reviews, and trailers of the advertised movies.

A movie's advertisement poster is just one example of a physical object that can serve as an anchor for digital content. For almost all products and services that are advertised with posters there is a digital counterpart, such as a website, available. We can therefore assume that most companies want to guide potential customers to this web page [PHN+08]. Furthermore, Google Goggles and similar applications are able to recognize a large number of other objects including music CDs, book covers, magazine covers, famous artwork, logos, sights, and even wine labels. Related digital content is available for almost all these physical objects and virtually all physical media has a digital counterpart.



Where the Wild Things Are
Movie

wild things are movie
Similar Image

Abc JELL SOON
Text

Google goggles labs

Movie
Where the Wild Things Are

http://www.goog...

Max Records , Catherine Keener , James Gandolfini , Paul Dano , Catherine O'Hara , Forest Whitaker , Michael Berry Jr. , Chris Cooper , Lauren Ambrose
Rating: 8
Date: 2009
length: 104

Product Search

Web Options ▾

Where the Wild Things Are
Official Warner Brothers site offers synopsis, cast, filmmakers, gallery and videos. Includes iPhone app, ticket and show ...
wherethewildthingsare.warnerbro... - Options ▾

Where the Wild Things Are (2009) - IMDb
Directed by Spike Jonze. With Max Records

Figure 1.2: Screenshots of Google Goggle after taking a photo of a movie's advertisement poster (left) and the view after selecting one of the results (right).

Existing applications focus on the interaction with physical entities that are of public interest. While not implemented today other use cases are also possible. The technology used to recognize book covers and magazine covers can also be used to recognize personal physical media. Printed photos, for example, could be connected with their digital counterpart on a photo sharing website. Thereby, it would become possible to retrieve a description, comments, and other information that cannot be directly provided by the physical photo. This would enable to maintain the tangibility of the printed photo merged with the interactivity, personalization, and real-time features of digital content.

1.2 Challenges and Contribution

The numerous commercial applications and earlier research prototypes show that the algorithms to recognize a broad range of physical objects are available today. The availability of these applications and the interest of a large number of users prove not only the commercial interest but also the high potential. When looking at the scenario from a human-computer interaction perspective, however, the scenario also shows that there is room for improvement. The user must explicitly trigger the recognition by taking a photo of an object. This input is decoupled from the provided feedback and it takes up to several seconds before the feedback is provided. The decoupled feedback is an immanent characteristic of the underlying interaction technique and related techniques might therefore provide a higher usability. Even worse, in the described scenario, it is not clear which objects can be recognized by the system. As a result the user has to use a trial and error strategy to determine which objects can provide digital content even without the support through direct feedback. Furthermore, when developing the interface design of such applications one cannot rely on previous research.

Previous work focuses either on users' performance when executing abstract tasks or investigates very specific use cases. In the context of mobile camera-based interaction we therefore face the following central challenges:

(a) Camera-based interaction techniques have not been systematically compared with each other and traditional interaction techniques. It is not clear which technique provides the highest usability and user experience. Therefore, designers and developers do not know which technique to select when developing mobile applications.

(b) No specific design principles for mobile camera-based interfaces to interact with physical objects have been developed. Developers can therefore not take advantage of empirically verified principles when designing the interface.

(c) Mobile camera-based applications use the device's screen as a peephole to the real world. Due to the limited screen size it is inefficient and can even be ineffective to determine which objects one can interact with. Means to present the availability and location of physical objects that can serve as anchor for digital information are required but have not been investigated.

Facing these challenges we examine camera-based interaction techniques for accessing information connected to physical objects with a focus on the human factors. Therefore, the work is based on a series of user studies and rich user feedback. To collect meaningful feedback from potential users it is required to confront them with intelligible and realistic scenarios. Therefore, each study needs to investigate the interaction techniques for a concrete use case using a prototype that addresses a specific scenario. We hereby ensure that participants understand how the interaction technique would be used in their everyday life. As we investigate novel interaction techniques it is not possible to study those using existing systems. Therefore, it is required to design and implement robust prototypes to test the interaction techniques with potential users. To make the findings generalizable beyond a specific use case, we consider five different use cases that we investigate through user studies. We developed prototypes to interact with posters, printed photobooks, paper maps, sights, and music CDs. In the strict sense each individual study cannot be generalized beyond the respective use case. Findings from developing and evaluation a user interface to interact with music CDs, for example, might not be applicable to interfaces for other types of physical objects. Addressing different types of physical objects, using different methods when conducting the studies, and being conservative when interpreting results enables to generalize our findings beyond the scope of a specific study. Furthermore, we augment our results with other researchers' findings which enables to derive general conclusions for a class of physical objects.

We systematically analyse potential interaction techniques described in the related work to identify classes of camera-based techniques to interact with physical objects. (a) Through four consecutive user studies we compare the identified camera-based interaction techniques with traditional approaches to show their advantage and compare different camera-based techniques to determine their differences. Thereby, we identify the most usable and least demanding camera-based interaction technique. For the identified interaction technique we systematically investigate the interface design for physical objects that are in the user's current focus as well as the interface design for objects that are beyond the user's current focus. (b) Through participatory design we explore the design space of the user interface to interact with objects in the user's current focus. Potential interface designs are compared through controlled experiments to derive validated user interface design principles. (c) To also visualize the availability and location of objects that are beyond the user's current focus we revise recent approaches to visualize off-screen objects through controlled experiments. We iteratively refine and extend existing approaches and apply them to the interface design of camera-based mobile interaction techniques. In total we conducted 15 studies with up to 3,934 participants using participatory studies, controlled experiments, and explorative studies both in the field and in the lab.

Figure 1.3 provides an overview about studies that are described in this thesis. The work is structured by three steps. We first investigate potential interaction techniques for physical objects, then study the on-screen interface design and finally study the off-screen interface design. In the following we outline the three challenges addressed in this thesis and describe our contribution to meet them.

(a) Comparison of Techniques for Mobile Interaction with Physical Objects

While the user receives the desired information in the scenario described in 1.1, the interaction between the user, his or her mobile device, and the advertisement poster might leave room for improvement. In the scenario an interaction technique that we call Point & Shoot is used to explicitly select the poster. This interaction technique is used by commercial applications today. Previous research has not shown, however, if this interaction technique provides a real benefit compared to established interaction techniques, such as using a text-based search engine or typing an URL using a soft keyboard in a Browser's address bar. Furthermore, Point & Shoot is just one camera-based interaction technique. It can be argued that Point & Shoot might not fully comply with fundamental guidelines for user interfaces e.g. supporting the efficiency of the user by avoiding that the user must wait for the system [Nie94, Tog03], user control and freedom [Nie94] and the demand for explorable interfaces [Tog03]. Furthermore, other camera-based interaction, such as Continuous Pointing and especially handheld AR investigated in this thesis might be superior.

We analyse three different camera-based interaction techniques for physical objects in four subsequent user studies. Using a controlled experiment we compare the camera-based interaction Point & Shoot with manual text entry [PHN^{+}08]. Based on this eval-

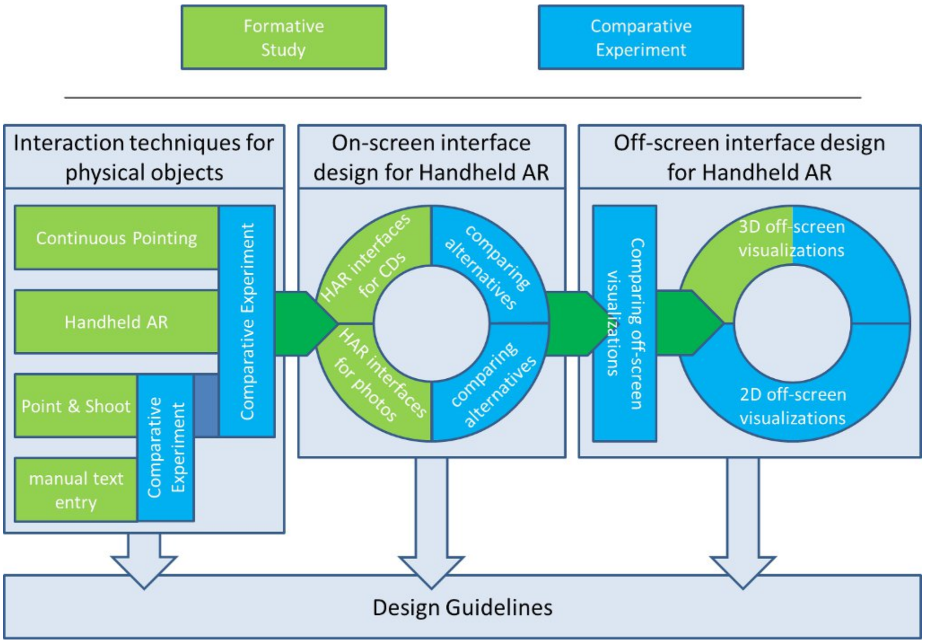

Formative Study
Comparative Experiment
Interaction techniques for physical objects
On-screen interface design for Handheld AR
Off-screen interface design for Handheld AR
Continuous Pointing
Handheld AR
Point & Shoot
manual text entry
Comparative Experiment
Comparative Experiment
HAR interfaces for CDs
comparing alternatives
HAR interfaces for photos
comparing alternatives
Comparing off-screen visualizations
3D off-screen visualizations
2D off-screen visualizations
Design Guidelines

Figure 1.3: Structure of the studies that are part of the thesis. Formative studies explore the design alternatives while controlled experiments compare different design alternatives.

uation we show that Point & Shoot outperforms the soft keyboard-based interaction in terms of speed and user satisfaction. Based on explorative user studies we show that current algorithms are sufficient for using Point & Shoot under realistic conditions [HB08]. Our results indicate that Point & Shoot is not only usable by young early adopters but also by elderly users without a technical background. Interaction with posters using the interaction technique Continuous Pointing is further analysed in an explorative user study [HSB09]. Analysing three camera-based interaction techniques in a controlled experiment we provide evidence handheld AR is significantly more valued by participants compared to the other interaction techniques when accessing information about physical objects [HB12]. In conjunction, the conducted studies do not only show that handheld AR is the superior camera-based interaction technique but also show that it is better suited than manual text entry for the considered task.

(b) Design Principles for On-Screen Content and Controls in Handheld Augmented Reality

Selecting an appropriate interaction technique is important but the design of the user interface is crucial when developing interactive systems. This, of course, also applies to applications that are based on a camera-based interaction technique. However, there is no guidance for developers that design the interface for such emerging applications. In particular, handheld AR is a young field. Developing applications or prototypes requires the use of architectures and algorithms that are currently subject of intense discussion and research. Therefore, handheld AR research is currently dominated by technical development and only the basic characteristics of handheld AR interaction have been studied using abstract tasks. The interface design, however, has been mostly neglected so far.

As we determined that handheld AR should be the preferred camera-based interaction technique we investigate the interface design of handheld AR application [HB10a, HB11c]. Addressing two different use cases and different types of studies enables to assume that the results can be transferred to other application domains beyond the particular types of physical objects used throughout the studies. We show that information connected to a physical object should be displayed in the reference system of the object. In contrast, we found that input controls that must be touched with the finger should not be displayed in the reference system of the object but aligned to the display remaining at a fixed position. Furthermore, we determine that highlighting physical objects inside the camera image is demanded and we provide guidance for doing this. We show that greying out the background and displaying only the objects with colours is well received and can easily be transferred to use cases where small and medium size objects are augmented. Based on our observations we can conclude that the interface of handheld AR applications should be in landscape or at least provide the option to switch to a landscape mode. Finally we found that current algorithms for handheld AR still leave room for improvement and conclude that effort invested in improving the algorithms is actually worth the hassle.

(c) Interface Design for Off-Screen Visualization in Handheld Augmented Reality

In order to interact with physical objects using any camera-based interaction technique the user must know that augmentable physical objects exists nearby. Furthermore, the user must determine the location of these objects. Considering the exemplary poster that advertises a movie it is not obvious for the user that this poster can be used as an anchor to related content nor is it obvious that there are other posters nearby that also carry information. Because only a small fraction of physical objects is connected with additional information today and presumably also in the near future the objects lack the desirable affordance. Assuming that physical objects cannot be reasonably altered the mobile device must support the user in identifying nearby augmentable objects.

Visualizing the location of augmentable objects is crucial for handheld AR application as long as the objects themselves lack affordance. To base our work on solid ground we revise previous work about off-screen visualizations for digital maps [HB10c, HPB10]. In two studies we show that these off-screen visualization techniques for digital maps scale differently. While the visualization technique Halos performs better for a low number of objects, arrow-based visualizations outperform Halo for a larger number of objects. Conducting the studies by publishing prototypes in a mobile application store and attracting more than 5,000 participants from all over the world the studies do not only allow strong conclusions about off-screen visualizations but also show that conducting controlled experiments using mobile applications stores is possible. Based on the results an off-screen visualization for using handheld AR with printed maps is developed and evaluated [HB10b]. We show that an off-screen visualization not only reduces the perceived task load but also reduces the task completion time. Transferring off-screen visualization to applications for interacting in 3D we design and evaluate an arrow-based visualization based on a comparison of three alternatives [SHB10]. We show that the developed visualization outperforms a mini-map and could improve the usability of current commercial applications.

1.3 Outline

After introducing the topic and outlining the contribution of this thesis in Chapter 1, Chapter 2 classifies the related work. The current state of the art is reviewed and a discussion of different approaches to connect physical objects with digital information using mobile devices as a mediator is provided.

Chapter 3 presents an analysis of camera-based interaction techniques to access information about physical objects by describing four consecutive user studies. After providing a description of the addressed interaction techniques the differences between manual text input and the camera-based interaction technique Point & Shoot are investigated using a controlled experiment. The characteristics of the interaction techniques Point & Shoot and Continuous Pointing are further analysed in two explorative user studies. The addressed camera-based interaction techniques Point & Shoot, Continuous Pointing, and handheld AR are finally compared in a controlled experiment. We close this chapter with a summary and an outline of the implications of our findings on the design of mobile applications that are used to access information about physical objects.

In Chapter 4 the interface design of handheld AR systems is investigated. Printed photo books and physical CDs are used as exemplary types of physical media. Design solutions to augment both media types are explored using a participatory approach. Design solutions proposed by participants of two user studies are consolidated and implemented as software prototypes. In two experiments the resulting interface designs are compared to determine their usability. We show that an object-aligned augmentation is more efficient for information presentation. In contrast, it is also shown that input controls should not be presented via an object-aligned presentation. We close the chapter with a summary and an outline of the implications of our findings on the design of prospective handheld AR applications.

Chapter 5, investigates how off-screen visualizations can be applied to handheld AR. Existing off-screen visualizations for digital maps are revised using a realistic task and a very large number of users to find a starting point for developing off-screen visualizations for handheld AR. It is confirmed that arrow-based visualization outperform the circular approach Halo and it is also shows that it is possible to conduct user studies by publishing an application in a publicly available mobile application store. In the subsequent study, the effect of off-screen visualizations on the interaction with physical maps is investigated. It is shown that an arrow-based off-screen visualization significantly improves the users' performance. Finally, off-screen visualizations for 3D objects are investigated by developing and comparing three visualization techniques for highlighting POIs. Based on the results the visualization's design is revised and compare with a baseline in a controlled experiment. It is shows that the developed off-screen visualization outperforms the commonly used mini-map for handheld AR applications.

Chapter 6 concludes the work with a summary of the thesis, an overview about our contributions, guidelines for developing camera-based mobile applications, and an outlook to future work in this field.

Publications

Excerpts of this thesis have been published in scientific journals, conferences, and workshops: [HB12], [HB11c], [HPP+11], [HB10b], [SHB10], [HB10a], [HPB10], [HB10c], [HHB10], [HLB+10], [HSB09], [RZHR09], [PHN+08], [HRR+08], [HB08], [HRL+08], [SPHB08], and [HLL+07].

A number of further publications beyond the specific scope of this thesis have been published by the author: [HRB12], [Hen12a], [HP12], [Hen12b], [PHF+12], [AH12], [MFP+12], [HRB11], [LHP+11], [HB11b], [HB11a], [PHB11], [PPHB10], [ERZHR10], [PHB09], [PHHB08], [HHPB08], [HHB07a], [PHB08], [HHB07b], [PHHB07], and [HHB06].

Several bachelor theses, master theses, and diploma theses have been co-supervised by the author and served as basis for many results described in this thesis, most notably [Lan11], [Blu10], [Löc10], [Sch09], [Blu09], [Kön09], [Pop08], [Naf08], [Pop07], [Pie07], and [Nüs07].

2 Related Work

Handheld devices became an integral part of our life. In 2009 a survey showed that three quarters of the respondents never leave home without their phones and more than a third even stated that they cannot live without their mobile [Syn09]. The usage of mobile devices is not only about calling and texting anymore but more and more about browsing the web, writing emails, playing games, consuming media, using location-based services and participating in social networks [BHSB11]. An upcoming, very important aspect is that the mobile device is conceptually not just used for the interaction with digital information but also for the interaction with the real world we are actually living in [BBRS06, RPF+06, BDLR+07, HBR+08, ZHRR09]. Influential work in this area, for example, showed that "palmtop computers" can be used to access information about physical objects [Fit93], introduced concepts for mobile interaction with people, places and things in the real world [KBM+02], and more recently coined the notion of "Mobile Interaction with the Real World" [RPF+06, BDLR+07, HBR+08, ZHRR09].

Considering the development towards location-based mobile services in the last years (see [Küp05, JW08]) it can be assumed that this area is now very well explored and intensively applied in practice. The indirect mobile interaction with other users who are not co-located is also very much explored through application areas like voice communication, text messaging, social networking applications, and instant messengers. In contrast, the direct mobile interaction with co-located users and the interaction with the surrounding environment is still a field with many open research questions and issues. It is very difficult to use a mobile phone for establishing a connection with another person's device or get information about nearby physical objects. Although it is technically possible via technologies like Bluetooth, WiFi or Near Field Communication (NFC) most people do not use it because of complicated technical device discovery and selection processes. The main issue is that available technology does not support this interaction on a level we are used to. While the interaction takes place in the physical realms we cannot rely on the means of interaction we are used to use in the real world. Today, natural interaction techniques like touching and pointing are not available in mobile interaction.

A significant corpus of research aims at extending the interaction between a person and a handheld device towards using handheld devices for the interaction with the surrounding world. In this survey we call this strand of research 'Mobile Interaction with the Real World' and provide an overview about research and development in this area. In the next section we define the scope and provide a classification that distinguishes between touch-based interaction, Point & Shoot interfaces, Continuous Pointing and handheld AR. Along these four categories we give an overview of the approaches and discuss their characteristics. We summarize the findings and provide an outlook that identifies open challenges for the field.

2.1 Overview and Classification

Already in 1968 Sutherland presented his pioneering research on head-mounted displays (HMDs) [Sut68]. His work exploited the "kinetic depth effect" that occur if an image presented by a two-dimensional display changes in exactly the same way that the image of a real object would change for similar motions of the user's head. This basic effect has been the driving force behind "Virtual Reality" and the immersion in virtual environments. The base idea was further enhanced to "augment" the user's visual field with additional information registered to the surrounding enabling an "Augmented Reality". In parallel to Caudell and Mizell [CM92] who assumedly coined the term Augmented Reality (AR) a number of groups presented early work on AR in the nineties. Feiner et al., for example, presented "Knowledge based Augmented Reality for Maintenance Assistance" (KARMA) [FMS92, FMHS93, FMS93] and Bajura et al. used AR for visualizing medical data during patient treatment [BFO92, OBF94].

While most work on AR in the early nineties focussed on using HMDs, Fitzmaurice presented the vision to use "palmtop computers" for virtual reality [FZC93] and AR [Fit93]. He described for instance an application with which the user can point at locations on a physical map with a mobile device to get additional information about points of interest. Through user studies Fitzmaurice showed that a registered handheld display that can be moved in space provides a similar perception of depth than a large static screen [FZC93]. Fitzmaurice's work was the starting point of research that investigates the uses of handheld devices to provide information about the surrounding physical world. After almost twenty years of research and development a number of commercial applications put Fitzmorice's idea into practice: SnapTell's mobile application, for example, enables to take a photo of products to receive information about it. Germany's national railway company enables users to buy a ticket by touching so called Touchpoint with an NFC enabled phone. A number of handheld AR applications exploit the location and orientation sensors of current smartphones to display points of interest.

While research in this domain, especially in the early years, focussed on investigating technological solutions, the commercial success of current applications and the expected spread in the next years makes it crucial to consider the interaction design and the human factor when developing such applications.

2.1.1 Definition and Scope

A broad range of research investigates the interplay between users, handheld devices, and objects in the real world. Rekimoto and Nagao call this interaction "Augmented Interaction" [RN95] and in their earlier work Rukzio et al. calls it "physical mobile interaction" [Ruk06, RLS07]. In the line with the later work by Rukzio et al. [BHP$^+$08a], a workshop series [RPF$^+$06, BDLR$^+$07, HBR$^+$08, ZHRR09] that presumably coined the term, a recent special issue with this title[1], and an upcoming special issue with the same title[2] we call this research area "Mobile Interaction with the Real World" (MIRW) and define it as follows:

> Mobile Interaction with the Real World is the research field that investigates the interplay between users and physical objects in the proximity using handheld devices as mediator for the interaction.

MIRW originates from the same roots as Ubiquitous Computing (UbiComp) [Wei02] and Tangible User Interfaces (TUIs) [IU97, UI00]. It is similarly driven by the search for post-desktop techniques for human-computer interaction. MIRW is different from TUIs and UbiComp because it explicitly considers a handheld device that mediates between the user and the real world. In contrast, Weisser explicitly states that ubiquitous computing "will not require that you carry around a PDA" [Wei02]. Thus, MIRW might be seen as a temporary solution towards ubiquitous computing. Explicitly considering a user's personal device can, however, help to address practical issues of ubiquitous computing (e.g. privacy, surveillance, and energy consumption) because personal mobile devices provide private storage, processing power, and can steers the interaction. MIRW is also closely related to location-based services (LBS - see [Küp05, JW08] for an overview). LBS provide the user with services based on his or her location while MIRW provide the user with services based on the relative position of nearby objects. Similarly we do not consider work on the interaction with public displays (e.g. [CDF$^+$05, BRSB05]) or handheld projector-based interactions (e.g. [CB06, HRG08]) as both interaction techniques focus on additional displays that partially replace the handheld device.

Over the years the technologies to develop MIRW applications and prototypes changed and will assumedly develop further in the future. The findings about interaction techniques that researchers proposed, implemented, and tested with users, however, provide guidance for current and future research and development.

2.1.2 Classification

In MIRW systems the mobile device mediates the interaction between the user and nearby physical objects. Conceptually the mobile device senses the presence of physical objects and provides feedback to the user. MIRW systems can accordingly be classified by the way objects are sensed and when the system provides feedback.

The interaction technique is tightly couples with the technique used to sense physical objects in the user's surrounding. Interaction techniques can be classified along the

[1] Special issue on Mobile Interaction with the Real World in the International Journal of Mobile Human Computer Interaction: http://www.igi-global.com/Bookstore/TitleDetails.aspx?TitleId=45623

[2] Special issue on Mobile Interaction with the Real World in the Journal of Pervasive and Mobile Computing: http://www.elsevierscitech.com/cfp/cfp_mobile_interaction_with_the_real_world.pdf

number of dimension in which they provide information about objects. In the case with the lowest degree of freedom the physical object has to be touched by a mobile device. An object is either touched and thus "connected" or not touched and not connected. This interaction is conceptually binary and thus one dimensional. A typical example is the use of NFC tags attached to an object or part of an object in conjunction with an NFC reader integrated in a mobile phone. In another one-dimensional case the user points at objects using a mobile device. Objects are sensed along a straight line. A real or virtual beam is send out from a mobile device to determine the object that is hit by this beam. Consequently, an object is either hit by the beam or not - also resulting in a conceptually binary and one dimensional technique. Välkkynen and Tuomisto for example use light sensors on posters and other object that can be illuminated with a light beam emitted by a PDA [VT05]. Different are Point & Shoot systems that recognize objects in a cone, for example, sensed by a mobile device's camera. The camera projects the scene in front of the camera on a 2D image that is analyzed. An example is the use of 1D Barcodes or QR Codes to provide product information. Handheld AR applications sense objects in almost the same manner. Mobile phones' displays serve as peepholes in the augmented space. Point & Shoot as well as handheld AR applications both recognize physical objects anywhere in the two-dimensional image (apart from technical limitations). The differentiation between one-dimensional and two-dimension techniques forms the y-axis of the diagram shown in Figure 2.1.

While the sensing techniques also constrains the way the system can react, conceptually, MIRW interaction techniques can be distinguished in techniques that react to the user's discrete and explicit action to provide feedback and interaction techniques that provide continuous feedback. Examples for reacting to a user's discrete action are systems where the user touches a NFC-equipped object with a NFC reader. While a NFC-based system might sense object over a distance, the system usually only reacts if the reader touches the object. Similarly, so called Point & Shoot application require an discrete trigger. The user takes a photo of an object which is analyzed to provide the user with information about the object. The user provides the trigger by taking a photo and the system provides according feedback. In contrast, continuous feedback is provided by systems that enables users to point at physical objects, e.g. with a continuous light beam, to retrieve information. Likewise, continuous feedback is provided by handheld AR systems. The system permanently augments the camera image with additional information and thereby provides feedback without an explicit action by the user regarding a specific object. The differentiation between discrete and continuous feedback is reflected in the x-axis of the diagram shown in Figure 2.1.

The classification of MIRW interaction techniques is similar to the schema of smartphone-based interaction techniques for ubiquitous computing applications provided by Ballagas et al. [BBRS06]. Our classification concretize and simplifies the schema for MIRW applications. In the following related work is discussed along the four identified categories touch-based interaction, Point & Shoot interfaces, Continuous Pointing and handheld AR. Afterwards, a summary of the findings and identified open challenges are provided.

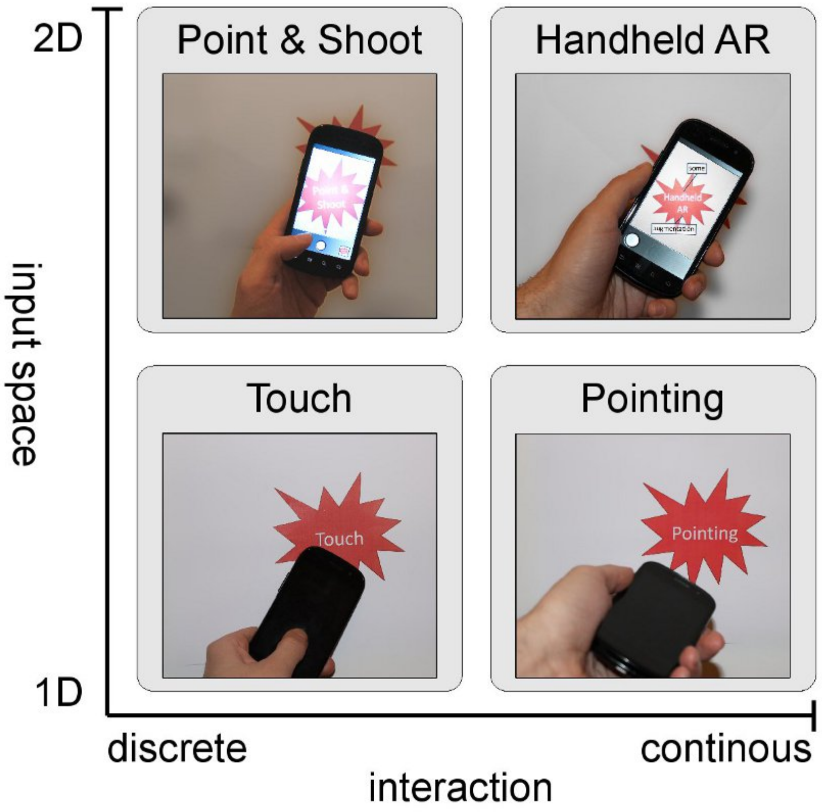


Figure 2.1: Classification of interaction techniques by the degrees of freedom used to sense and visualize physical objects.

2.2 Touching and Hovering

Touching with the fingers is an interaction modality that has been used even before computers and digital devices became available. Hardware buttons are likely the most pervasive input techniques to control electronic devices in the form of power switches, hardware keyboards and mouse buttons but are also used to control analogue devices such as light sources and most electronic devices. Using mobile devices to touch objects can be used to facilitate a similar interaction modality. Instead of touching an object's surface with the finger, one touches it with a handheld device. Using the mobile device as a mediator enables to make objects interactive that are not equipped with means to detect touches. The mobile device can further be used to store information and steer the interaction flow.

Early work beyond the use of an input device for direct manipulation of stationary computer (such as light pens) has been done by Rekimoto [Rek97] in 1997. He designed an interaction and implemented an according basic prototype to transfer data from one visual screen to another by first touching the representative of a data item on one screen with a "pen" and virtually releasing the data item by touching with the pen on a second screen. Rekimoto also describes a prototype where paper documents can be used to pick-up data with the pen. Rekimoto's work is exemplary for HCI research that anticipated the technical scope. Want et al. [WFGH99] identified limitations of research at that time and proposed to go beyond highly specific and expensive prototypes. They propose RFID tags in conjunction with handheld computers to connect everyday objects with digital information in a lightweight and low-cost way. Using RFID they implemented an according prototype. RFID and its extension Near Field Communication became the most widely used technology used by research and industry for touch-based MIRW interaction in the following years. In the following we provide an overview about touch- and hovering-based interaction, from simply retrieving and adding information to the interaction with complex services, along the complexity of the exchanged information.

2.2.1 Retrieve and Add Information

Nath et al. identified a number of use cases for mobile phones with an integrated RFID reader [NRW06]. In particular, they propose that products equipped with an RFID tag could provide product-specific information on the phones display, for example, nutritional information for food products.

Korhonen et al. presented an alternative architecture for such a system that pushes websites to the user's phone if the phone touches an object [KOKV06]. Unlike other implementation they installed fixed RFID reader in the environment and equip mobile phones with RFID tags. In two user studies Korhonen et al. evaluated their system and conclude that the interaction was found as an easy way to access location-based mobile web sites. As one of their main findings they highlight that the readers (i.e. the objects that provide the physical hyperlink) should have a prominent appearance.

Garner et al. extended the idea to retrieve information by touching objects by enabling to also add additional content [GRCE06]. They presented the Mobile SprayCan System that enables to add virtual graffiti tags to the environment, using mobile phones and RFID tags. SprayCan "site markers", small plastic cards, are added to the environment and users can add their graffiti tag to the site marker by touching it with their phone. The virtual graffiti tag is represented by small individual images. From the informal user study the authors conclude that the tagging process was perceived to be extremely easy. Most of the participants commented that the act of actually touching the object you wished to communicate with seemed very obvious and natural.

Product rating and recommendation is probably the most common use case that requires adding and retrieving information proposed by researchers. Assuming that virtually all products will be equipped with RFID tags von Reischach et al., for example, developed Apriori [vRM08, VRGMF09] a system that allows users to receive and submit product ratings by touching a product. Regarding the interaction design they state that "the system needs to be interaction-wise extremely simple and straightforward" based on a formative user study.

Mäkelä et al. [MBGH07] conducted a study with RFID-tags and visual markers to investigate the usability and acceptability of these technologies. The authors showed that most participants did not know how to trigger the interaction with them. They suggest that the minimal visual interaction cues caused misconceptions and usability problems while interacting with the tags. The tags were assumed to contain direct information in encrypted form in contrast to acting as references to networked data resources. Arnall states that in order to retrieve information from a physical objects the user must be aware that the object has digital function, information or history beyond its physical form [Arn06]. As NFC and RFID tags are designed to be embedded into objects they do not have a predefined form that enables the user to identify them. Arnall explores the design space to visualize that a physical thing can provide a link to digital information by highlighting existing icons for RFID-based touch interaction and presenting a number of additional sketches of icons [Arn06].

Touch-based interaction using RFID or NFC to simply retrieve or add information for individual physical object has been described in a large number of further work (e.g. [SV07, GBM08, SRRP08]). Probably because of its simplicity the interaction technique itself has, however, received little attention. The main conclusion from previous work is that the interaction technique is easy to use [KOKV06, GRCE06, GBM08, VRGMF09]. Embedding NFC tags in objects allows interactions without changing the object's visual appearance. A crucial factor is therefore, that the user can easily determine that the object has digital function, information or history beyond its physical form [KOKV06, Arn06, MBGH07]. Välkkynen et al. proposed to use a consistent visualizations for physical hyperlinks [VTK06]. Comparisons of different visual representations, however, mainly addressed more complex use cases (see below).

2.2.2 Touching Services

While retrieving and adding information by simply touching a physical object with a mobile device has been explored by researchers, invoking services using this interaction technique has gained considerably more attention. In particular, mobile ticketing and mobile payment have been a driving force for developing NFC [For08]. Zmijewska, for example, reviewed different approaches for mobile payment and highlight the ease of use of NFC-based approaches because they are "employing the natural human behaviour of touch" [Zmi05].

Ghìron et al. [GSMM09] developed an NFC-based Virtual Ticketing application that enables to buy bus tickets by touching an NFC equipped poster with a mobile phone and also to pass a ticket to other persons by touching their phone. Based on a formative usability study the authors describe a number of observations. Regarding the interaction design they observed that participants fear to buy more than one ticket by mistake and propose that the phone should alert the user if a ticket is already available. The authors highlight that all participants appreciated the idea to provide some of their tickets to friends or relatives, having an NFC phone, by means of a simple touch. Furthermore, they found that the used poster was difficult to understand mainly because of the NFC tag's labelling.

Cappiello et al. [CPV09] developed a touch-based remote grocery shopping prototype also using NFC tags and an NFC equipped phone. The user touches groceries or an RFID tagged picture of a product at home with a phone in order to create a shopping list. By touching a "buy icon" the order is send to a grocery store that delivers the products. The authors' user study shows that often users need multiple attempts to select a product by touching it. Comparing the touch-based prototype with a shopping process using a minimalistic web page showed that participants need more time using the touch-based approach. The authors, however, mention that the comparison was not entirely fair because of the web page's simplicity. A similar system has been implemented and evaluated in an eight-week study by Häikiö et al. [HWI+07]. Instead of touching products directly elderly participants could order a meal by touching a paper menu attached to an NFC equipped plastic stand. The authors suggest adding additional auditory feedback if an NFC tag has been read to improve the usability. They further assume that the user's attitude towards new technology is a crucial factor for services similar to the one used in the trial.

Geven et al. [GSF+07] investigated different use cases in five complementary studies to analyse how novice users interact with NFC-equipped mobile devices and how their experiences change when using NFC more often. They show that novice users often do not know which part of an object has to be touched and are unsure how to align the mobile device. Furthermore, the authors state that the sequence of interaction was unclear. Participants had problems building a mental model of interaction to successfully complete tasks. The authors recommend that the spot that should be touched and the position of the phone's NFC antenna should be clearly marked, that the interaction process should be shown with a step-by-step description, and that critical processes (e.g. payment) should require an additional confirmation. Similar use cases have be investigated by Falke et al. [FRD+07] but in a controlled usability study. While the authors highlight the interaction's good usability the described issues are consistent with Geven et al. [GSF+07]. Participants had difficulties to successfully touch a touch point because they over-estimated the scanning range, under-estimated the time the system needs to scan a tag, and did not know which part of the phone must touch the object. Furthermore, it was not clear to the participants what kind of service is associated with particular objects.

Most studies that investigate touch-based interaction using a handheld device are conducted in a highly controlled environment a very short time frame. An exception is the study conducted by Hardy et al. who investigated their MyState system [HRHW11]. The system allows users to equip objects and locations with NFC tags. By touching these tags the user shares self-defined status updates on a social network. The two conducted studies last for several weeks. Hardy et al. observed both personal use, such as retracing steps and activity history, and social use, such as synchronizing activities, expressing moods, games, and tracking shared items. The authors argue that users chose to continue using the prototype even if no incentive was provided.

Accessing services by touching physical objects with a mobile phone offers a large number of possibilities and the implementation of various use cases. Technological improvement might help to overcome simple usability issues such as small scanning ranges and long delays. From the perspective of the interaction design the most severe problem is that it is not clear what is associated with a particular object or touch point [FRD+07, GSF+07]. More generally the interaction process might not be clear and consistent. Different solutions have been proposed to tackle this challenge. Välkkynen et al. proposed to use a consistent visualizations across different use cases [VTK06]. Anokwa et al. also recognizes this lack of an interaction model [ABPW07] and propose that the mobile phone should take on the properties of existing objects. E.g. by touching a movie poster the user is offered a list of actions that initiate the download of virtual objects that represent the analogue of a physical object like a movie ticket. The phone "transforms" into the virtual object and the offered actions reflect what could be done with the physical object. The virtual movie ticket would, for example, offer to pass itself to another user.

2.2.3 Interacting with Complex Services

The work discussed so far uses touch-based interaction mainly to select content or invoke services. The physical object or an object's touch point is used as an anchor for an object rather than an achor to specific services connected to the object. The interaction process after the user's touch takes place on the mobile device. If the user, for example, touches an interactive poster different options are offered by a menu on the phone. In contrast, several researchers also investigated approaches to realize most of the interaction process with touch-based interaction.

A general framework for requesting services by touching different visual symbols has been proposed by Riekki et al. [RSA06]. They distinguish between NFC tags that are used as general symbols to identify the object a tag is attached to and special tags that are connected to specific actions. A printer might, for example, have one general tag accompanied by the three special tags: print, contact maintenance, and info. While all services are accessible through an object's general tag the special tags provide shortcuts to specific actions. The user study showed that special tags triggering services with one touch were preferred to the general tags that caused a list of alternatives to be shown on the phone's display. The participants, however, noticed that it might be difficult to find the correct symbol for a certain task among several adjacent special tags. Participants dismissed the idea to trigger services when a user enters the tag's proximity and preferred clear manual interaction because it gives a higher controllability.

Sanchez used multiple tags attached to an object to interactively control services [SRP08]. They attached labels with icons to control multimedia playback (e.g. next, play, and stop) and equipped them with RFID tags. By touching the labels with a phone the user controls a nearby multimedia player. A preliminary study suggests that the touch approach is easier to use than a UI on the phone that provides the same functions.

Broll et al. systemized the approach and distinguish between Single-tag interactions (STI) that use single tags as physical hyperlinks and keep the focus of interaction on mobile devices, multi-tag interactions (MTI) that map application features and UI elements to multiple tags on physical objects, and hybrid configurations that split features between tagged objects and mobile devices [BHP+08b, BHP+08a]. Broll et al. demonstrated the approach with "interactive" posters that are equipped with multiple locations that can be touched. Using such a poster the user can order movie ticket by selecting movie, cinema, time, and number of persons soley through touching different areas of the poster [BHP+08b, BHP+08a]. The authors also demonstrate cross-object interaction where users first select a tag (e.g. a cinema) on one object and "drop" the data on another object (e.g. a ticket service to buy a ticket to the cinema). Through user studies Broll et al. compared different configurations of traditional mobile UI elements and multi-tag interaction for a product order task [BH10]. The results suggest that the interaction should not be split between the phone and multi-tag interaction. Participants preferred multi-tag interaction except for more critical actions, like submitting an order.

Broll et al. and Hang et al. investigated how the design and accessibility of multi-tag interaction can be improved [BKHB09, HBW10]. In [BKHB09] Broll et al. compared different designs to improve the usability through guidance. They show that a dedicated start-tag that provides an explicit starting point improves the participants' performance compared to different visual cues on physical objects and the mobile device. The authors also discover the most common mistakes that occurred in mobile interaction with tagged objects. Hang et al. compared five different designs for an NFC equipped poster [HBW10]. The interaction was split between a mobile phone and touch-based interaction to different degrees. The authors conclude that the number of NFC-tags on the physical UI should be limited to design a clear UI and that the interaction should start witch touching an NFC-tag on the physical UI. They also argue that long-lasting interaction should take place on the phone while shorter interaction should use the physical UI.

Touch-based interaction using a mobile phone has also been used in conjunction with digital devices instead of static physical objects. Hardy et al. [HR08] equipped a projection screen with an array of NFC tags and Seewoonauth et al. equipped the reverse side of a Notebook display with tags [SRHH09]. Hardy's results suggest that the selection time of touching with a phone is comparable to a touch screen and faster than using the phone's joystick to control a cursor on a remote screen. In [SRHH09] Seewoonauth et al. compared three techniques to exchange pictures between a Notebook and a mobile phone. They used manual Bluetooth pairing and a standard user interface, touching a single tag attached to the Notebook to overcome the initial pairing, and an array of tags behind the Notebook's display to select specific pictures by touching the display with a phone. The conducted user study shows that participants prefer the touch-based interaction techniques and suggest that users prefer the technique that enables to touch the Notebook's displays to select content.

Little work tried to create a formal model of MTI. An exception is the work by Holleis et al. who proposed a Keystroke-Level Model (KLM) for MTI [HOHS07] based on a number of user studies. One aim of the model is to predict the user's timing in early stage of the development without actually performing a costly user study. Holleis et al. later refined the KLM [HSB11] to consider recent technical improvement of the underlying technologies.

2.2.4 Summary

Different technologies have been explored to build touch-based systems for interacting with physical objects. Early work used devices specifically adapted for this use case [Rek97] but already in 1999 researchers started to use RFID [WFGH99] and later NFC [Zmi05, For08]. The research interest in this interaction considerably increased from 2005 on. One reason is certainly major companies' commercial interest in NFC [OJ06] and the availability of standard hardware such as an NFC cover for the Nokia 3220 that has been used in a number of studies by different authors [RLC+06, LR06, RLFS07, RLS07, HWI+07]. It has been argued that one reason why large-scale deployment of NFC failed in most countries so far are the competing interests of the involved industrial stakeholders [OS10].

Different studies showed that there are use cases where touch-based interaction outperforms other approaches. [Zmi05] argues that NFC-based mobile ticketing is preferable from the users' perspective compared to other technologies including Bluetooth and infrared. Results from [MBGH07] suggest that users would prefer touch-based interaction using RFID compared to Point & Shoot using visual markers. Work by Sanchez et al. suggest that touch-based interaction might be preferred compared to a mobile phone UI [SRP08]. Similarily Broll et al. showed that there are users that prefer an ordering process using multi-tag interaction compared to a UI on a mobile phone [BH10]. However, [CPV09] showed that touch-based interaction is not necessarily more efficient compared to traditional solutions and that the technical implementation can highly affect the interaction.

While it has been identified that touch-based interaction can be more efficient than other interaction techniques, users have concerns to perform critical actions using this technique. [GSF+07] and [BH10], for example, showed that users clearly prefer to confirm payment using a traditional user interface presented on phones' displays. It seems unclear if this preference is the result of the interaction technique's novelty and the particular prototypical implementation or a fundamental characteristic. Another reoccurring finding is the need for adequate visual representations for actions that are executed when touching an objects or particular tags attached to objects [KOKV06, Arn06, MBGH07, FRD+07, GSF+07]. A number of proposals for visual representations have been developed [RSA06, VTK06, ABPW07]. Deploying the interaction technique in the large will, however, involve a number of stakeholders from competing companies. We therefore assume that finding a consistent visual language across specific implementations will be crucial for the success of touch-based interaction in the large.

2.3 Point and Shoot

Touch-based interaction requires having direct contact with the physical object the user aims to interact with. In contrast, the Point & Shoot interaction technique enables to triggers the recognition of a physical object from a distance. Still the user explicitly selects specific objects. This explicitness of Point & Shoot makes it similar to touch-based interaction. Just as users explicitly touch an NFC tag they can trigger to take a photo that provides an input for the mobile device. The interaction radius is, however, much broader and the user can interact with objects from a distance.

The technical bases of most work discussed in this section are visual markers that are recognized using an optical system. 1D barcodes can be seen as the first implementation of visual markers back in 1948. Ljungstrand et at. used 1D barcodes printed on paper to give bookmarks that link to web pages a physical form [LH99, LRH00]. Users can scan the barcode with a conventional barcode scanner attached to a computer to access the links. Since special barcode scanners are needed to read the barcode, conventional barcodes did not found their way into applications for common consumers. 2D barcodes [SSH97], such as the standardized Quick Response Code (QR Code) [Int06a], have been developed to increase the amount of encoded data and to allow robust decoding at high speed. Initially developed to be read by barcode scanner 1D and 2D barcodes can also be read by applications for mobile phones [RG04, OHH04].

2.3.1 Retrieve and Add Information

Early work that connects physical objects with digital content by pointing at the object and explictly triggering the recognition is "WebStickers" developed by Ljungstrand and Holmquist [LH99, LRH00]. WebStickers associate web pages with physical objects to use these objects as bookmarks to the World Wide Web. The developed system consists of printed barcodes that can be attached to post-it notes and other physical objects. Using a conventional barcode scanner connected to a computer, users can read the barcodes and access connected web pages. An informal evaluation of Webstickers suggest that users are positive about the system. Participants noted that such a system might be useful to exchange bookmarks in a group working together. Participants criticized the small reach of the desktop oriented system and propose a wireless reader that does not require bringing objects to the desk. In HP Labs' Cooltown project [KBM$^+$02] Kindberg accordingly extended the work, by enabling the use of mobile devices to access what they call physical hyperlinks [Kin02].

Since traditional 1D linear barcodes are designed to be read by manually operated laser scanner they are difficult to be recognized from a camera image [Ote99]. A further limitation is the small amount of data that can be stored [Ote99]. Therefore, different 2D barcodes, including Sony's CyberCode [RA00], Quick Response (QR) codes [Int06a], and Data Matrix codes [Int06b], have been developed over the years. Kato and Tan compared eight 2D barcodes for camera phone applications [KT07]. Since they conclude that no 2D barcode completely satisfied the criteria of a global 2D-barcode standard for mobile applications they suppose that "instead of choosing one standard 2D-barcode system, users might prefer different types of barcode systems according to the mobile applications they need" [KT07].

Brush et al. conducted a five weeks field study with a system that enabled participants to scan and annotate physical objects equipped with conventional 1D barcodes using a PDA connected to a barcode reader [BBCTSG05]. They observed a number of challenges including that participants found the external barcode scanner cumbersome and that some participants mentioned a fear of being seen scanning items in stores. Furthermore, Brush et al. observed a high level of frustration when items that participants scanned did not resolve or locating the barcode position on a product is difficult.

Klemmer et al. evaluated a system that connects books with digital audio and video content by using a PDA with an attached barcode reader that scans barcodes printed in the book. They found that while "there were usability problems, none of the users found the system conceptually difficult" and argue that "The concept of using paper transcripts as an interface to original recordings seemed perfectly 'natural'." [KGWL03]. Similarly, Boring et al. investigated the transfer of information from a public display to a mobile phone by taking a photo of icons presented on the display [BAB$^+$07]. They found that participants of the conducted study were able to successfully use the system after a short explanation and conclude that is an easy-to-use interaction technique. Boring et al. assume that the system is particularly compelling when the user is waiting e.g. for a train in a subway station.

Cheong et al. compared different visual markers and showed that participants prefer a visually appealing colour-based image code [CKH07]. A similar study that compares the usability of different visual markers has been conducted by Yoon et al. [YSY$^+$09]. Comparing pictograms, pictograms with an additional ColorCode barcode, and a pictogram overlaid by a ColorCode barcode they found that pictograms are preferred by participants. Consequently a number of authors [RSKH07, EAH08] including ourselves [HB08] argue that visual markers have inherent limitations. E.g. Rohs et al. state that "the markers obscure valuable map space" [RSKH07] and Erol et al. argue that "A significant issue for [marker-based] applications is the need to modify the document's format by introducing a machine-readable code that improves the utility of the document but disrupts its appearance." [EAH08].

Carter et al. compare different approaches, including different visual markers and content-based recognition, to link digital media with physical documents from a technical perspective [CLD$^+$10]. They conclude that content-based recognition using generic image features or word geometry features requires lower effort and cost for creating and inserting tags, support arbitrary user-defined tags, and enable a higher spatial density of tags on paper, compared to visual markers. Advantages of visual markers are the lower computational complexity and the option to encoded data in tags. Furthermore, Carter et al. consider the robustness and scalability of content-based approaches as not sufficient for commercial applications. Given the number of available commercial applications available today this finding can be considered as being out-dated.

Davies et al. studied the difference between two interaction techniques to acquire information about nearby POIs [DCDH05a]. They compared accessing information using Point & Shoot with an interactive list of POIs in the surrounding using a Wizzard-of-Oz prototype. Davies et al. conclude that "users appear happy to use image recognition even when this is a more complex, lengthy and error-prone process than traditional solutions" [DCDH05a].

Fan et al. [FXL$^+$05] as well as Tollmar et al. [TMN07] investigated Point & Shoot for mobile image-based web search. Participants could take photos of physical objects to trigger the search. Despite the lack of a statistical analysis the results suggest that image-based search could be faster than text-based search. Furthermore, Tollmar et al. conclude that Point & Shoot is perceived as more time efficient than text-based search even if the actual time to complete the tasks varied significantly. Tollmar et al. assumes that image-based searching is especially valuable if the search is about an unknown object.

2.3.2 Interacting with Services

Work that uses or investigates Point & Shoot for more complex use cases than just recognizing a physical object in order to accessing information connected to the object is rare. An exception is the work by Ballagas et al. who propose using Point & Shoot and an interaction technique the authors call "sweep" to interact with large public displays [BRSB05]. Using Point & Shoot users can select a position on the display. When the user shoots a photo of the display a grid of 2D barcodes is displayed on the large display. It is shown for target selection task that Point & Shoot results in a similar task completion time than using a phone's joystick. Point & Shoot, however, increased the error rate.

An explanation why Point & Shoot has rarely been investigated for interaction with complex services is provided by Rukzio et al. and Broll et al. [RLC$^+$06, BSR$^+$07, RLS07]. They compared Point & Shoot using 2D barcodes with an NFC-based touch interaction and a proximity-based approach. The investigated tasks included, for example, ordering a cinema ticket using a "smart" poster that required to subsequently selecting 2D barcodes using Point & Shoot. The results of a series of user studies show that the other interaction techniques are clearly preferred by users. The authors partially ascribe Point & Shoot's low performance to their specific implementation. The necessary concentration and physical skills needed to take a picture of a visual marker and, in particular, the lack of direct feedback seem like general characteristic of Point & Shoot.

2.3.3 Summary

Just as touch-based interaction, early work used different devices and approaches [KGWL03, BBCTSG05]. A number of later prototypes are based on visual markers [RLC$^+$06, BSR$^+$07, RLS07] in particular 2D barcodes that are recognized by a mobile phone. It has been argued that different types of visual marker have different advantages [KT07]. Studies showed that users prefer visually appealing markers [CKH07] and that pictograms are preferred compared to markers [YSY+09]. It has been argued that content-based object recognition is preferable [RSKH07, EAH08, HB08]. It has also been argued that content-based approaches are easier to maintain but the computational complexity might limit their use in commercial applications [CLD+10].

Studies showed use cases where Point & Shoot is faster than text-based search [FXL+05, TMN07] and that some users might prefer it compared to browsing a list [DCDH05a]. Compared to touch-based interaction it has been shown that Point & Shoot is not preferred by users but also argued that this might partially result from the performance of particular implementations. The lack of direct feedback and the required skills to take a photo might be inherent limitation of Point & Shoot [RLC+06, BSR+07, RLS07]. Point & Shoot shares the same need for a consistent visual language with touch-based interaction. Visual marker can, however, communicate the availability of an unspecified information or action but usually fail to communicate its type.

2.4 Continuous Pointing

Just as Point & Shoot, Continuous Pointing enables to interact with physical objects from a distance by pointing at them with a handheld device. Instead of requiring the user to explicitly trigger the interaction the handheld device continuously scans for objects in the direction it is pointing at. Camera-based Point & Shoot, for example, can be extended to Continuous Pointing by analysing the stream of images provided by the phone's camera and not just a single photo. Assuming a perfect object recognition algorithm, information about an object is shown by the phone as soon as an object appears in front of the camera. The user receives direct feedback and does not have to explicitly trigger the recognition. Camera-based Continuous Pointing can be implemented by either transmitting the stream of camera images to the server or by analysing the images directly on the phone (e.g. as we showed in [HSB09]).

2.4.1 Implementations and User Studies

To implement Continuous Pointing, Välkkynen et al. equipped a PDA with a laser pointer and [VT05] and instrumented the environment with sensors that detect the laser point. Rukzio et al. [RLC+06] used a similar system to compare the three interaction techniques touching, (Continuous) Pointing, and scanning (presenting a list of services provided by nearby objects). They conclude that participants preferred touch if the objects are near. If the objects to interact with a further away and there's a clean line of sight, participants preferred pointing. Only if all else fails participants prefer scanning. Reilly conducted a study that investigates the selection of objects by pointing a handheld pointer at physical objects without providing feedback [Rei11]. Factors that affects the pointing performance have been identified and Reilly concludes that, with the used technology, users are able to distinguish between four targets within a 180° range. It is argued, however, that incorporating additional factors (e.g. the device's pitch) could increase the number of distinguishable objects.

Few work falls in this category using the strict definition provided in Section 2.1.2. Most work studied systems that either require an explicit trigger (i.e. what we call Point & Shoot) or are registered in 3D (i.e. Augmented Reality). An exception is work that investigates non visual feedback for mobile interaction with the real world, in particular for persons with visual disabilities. Magnusson et al. evaluated a system for visually impaired users to use a mobile phone as a pointing device and displays the distance of objects within a certain angle using tactile feedback [MMRGS10]. They used GPS, magnetometer, and accelerometer integrated in recent smartphones to determine the phone's position and orientation relative to the position of physical objects' fixed position. Magnusson et al. further investigated the appropriate angle in which information should be conveyed [MRGS10]. They argue that 30° to 60° is appropriate if precise localisation is required but larger angles should be used if a low cognitive load is important.

2.4.2 Summary

Continuous Pointing prototypes have been developed using laser pointer attached to a handheld device [VT05, RLC+06] or using recent phones' inertial sensors [MRGS10, MMRGS10]. In contrast to other techniques Continuous Pointing has been used to specifically support persons with visual disabilities by using tactile displays. As the single vibrator for current phones prevent sophisticated tactile displays it is the only technique than can be used to provide continuous feedback.

Continuous feedback has not been compared to one of the other interaction techniques or to traditional interfaces. It could be argued that the direct feedback might provide an improved user experience compared to the explicit interaction provided by touch-based interfaces and Point & Shoot. The work by Reily [Rei11], however, suggest that feedback is required in order to support a reasonable number of objects to interact with.

2.5 Handheld Augmented Reality

While Continuous Pointing determines at which object the handheld device is pointing at, handheld AR registers the object's relative position in 3D. Recent progress in handheld AR is rooted in the technical progress of the AR community. Early work in the handheld AR domain used external sensors to track the position of the mobile device [Fit93, NIH01]. With the increasing computing power that became available in PDAs and mobile phones it became possible to implement marker-based AR on mobile devices. Wagner et al. developed a system that realizes handheld AR by recognizing and tracking visual markers on a PDA with up to 3.5 fps [WS03]. Similarly Möhring et al. presented the design and implementation of marker-based handheld AR using even less powerful mobile phones [MLB04]. With increasing processing power and refined algorithms, pose tracking from natural features [WRM$^+$08] and detection and tracking of multiple natural target [WSB09] became feasible on mobile phones. Despite the technical advancements that are expected to continue in the future [WS09] Billinghurst et al. stated that "much of the research in the field has been focused on the technology for providing the AR experience (such as tracking and display devices), rather than methods for allowing users to better interact with the virtual content being shown." [BKM09].

2.5.1 Devices for Handheld Augmented Reality

Different groups investigated the design of handheld AR devices. Grasset et al. describe the human-centred development of handheld AR binoculars [GDB07]. Through an experiment they show that the developed device is preferred by users compared to a HMD - at least for the short tasks they considered. Similarly Veas and Kruijff developed a handheld AR device based on an Ultra Mobile PC (UMPC) [VK08]. They state a UMPC without extended hardware can be sufficient for simple scenarios. In particular, if the interaction can be divided in short subtasks allowing the user to lower the device in between. For longer tasks and if using input controls is necessary the standard UMPC controls become tedious. Henrysson et al. developed and evaluated an application for manipulating virtual objects in an augmented scene [HBO05b]. They argue, while positioning virtual objects using the phone's motion is a valuable technique, rotating object should be better performed through a phone's keypad. In another study [HMB07] the authors come to similar conclusion. In addition, they observed that some users prefer holding the phone with both hands. It is argued that interfaces requiring the use of both hands might be difficult to operate.

Wither et al. investigated abstract selection and annotation tasks by comparing a HMD with using a handheld device either in a handheld AR configuration or in a tablet computer configuration [WDH07]. They conclude that using handheld AR may be more suitable than a HMD because throughout the studies handheld AR performed better or at least equal compared to the other conditions. Also focussing on low-level human-factors, Rohs et al. applied Fitts' Law [Fit54] to handheld AR [RO08]. They found that Fitts' law does not adequately model the users' performance when an object moves from outside of the camera image into the camera image. Rohs et. al adapted Fitts' Law accordingly [RO08] and also validated the model in the field [ROS11].

2.5.2 Interacting with Small Entities

The interaction design for different application scenarios using handheld AR has been adressed by researchers. Early attempts of handheld AR systems for gaming have been developed by Henrysson et al. [HBO05a] and Rohs et al. [Roh07]. Wagner et al. reported about their experience with "the invisible train" a collaborative game [WPLS05]. From a number of exhibitions they learned that using the system was so easy that visitors usually did not need any instructions, explored the game on their own, and explained the system to other visitors. Oda and Feiner developed an handheld AR game to investigate techniques to avoid that players physically block other players' view [OF09]. They show that transforming the 3D space in which the user moves their display can efficiently and unnoticeably be used to direct the display away from other displays.

One of the most frequently addressed application scenario for handheld AR investigated by HCI researchers is the interaction with physical maps. Rohs et al. compared different interaction techniques for map navigation with mobile devices [RSR$^+$07]. Participants of the conducted study had to find the cheapest parking lot on printed maps. Rohs et al. compared handheld AR over a paper map or over a grid of markers with a standard 2D mobile map controlled using the phone's joystick. It is shown that using the phone's joystick with a digital map results in a lower performance. The work is extended by an analysis of the impact of item density on handheld AR interfaces [RSS$^+$09]. Rohs et al. show that providing visual context through a physical map is most effective for sparsely distributed items and gets less helpful with increasing item density. Morrison et al. conducted a field trial to compare using handheld AR to augment a paper map with a standard digital map [MOP$^+$09]. One of their conclusion is that the "main potential of [handheld] AR maps lies in their use as a collaborative tool". Morrison et al. extended their work in [MML$^+$11] to analyse the effect of equipping teams either with a single mobile device or providing all team members with their own devices on collaborative interaction. One of their findings is that teams with multiple devices collaborate even if they do not have to. A recent survey about applications for augmenting maps, that also discusses handheld AR for maps, is provided by Schall et al. [SSPG11].

Handheld AR has also been used to augment physical media, such as paper documents. Liao et al. [LLLW10] developed and evaluated PACER that links paper documents to their digital version by using loose registration handheld AR. Different techniques to select parts of the document using touch gestures and phone motion are described. Based on a preliminary study the authors suggest that loose registration might reduce the demand for the participants because it does not require accurate phone-paper coordination. The authors also assume that one-handed operation may not work for all users. Users with small hands and short thumbs might have difficulty to operate the phone with only one hand.

2.5.3 Interacting with Large Entities

Another common application domain for handheld AR is the augmentation of larger entities such as buildings, sights, and squares. With the emerging of commercial applications, such as Layer and Wikitude, that exploit recent smartphones' GPS, magnetometer, and accelerometer researcher also adopted this approach. The advantage of this approach is that large amounts of already available geo-annotated data can be easily used. Karpischek et al. used these sensors to develop a handheld AR application to identify mountains [KMG$^+$09]. Tokusho et al. report from their experience when developing a handheld AR application that exploits the same sensors [TF09]. The main issues they faced are the inaccuracy of the phone's orientation and position sensors. In addition, they highlight the small field of view of the phone's camera that can require holding the phone in an awkward pose. Pombinho et al. implemented an application to search for POIs in the surrounding [PCAA10]. The preliminary user study suggests that providing a 2D overview in addition or as a replacement for a handheld AR interface might be valued by users.

As smartphones' location and orientation sensors have a low accuracy and systems based on these sensors cannot exploit an objects' shape the augmentation is only loosely aligned to the object. In contrast, computer vision-based approaches can determine the position and shape of an object insie the camera image more precisely. White and Feiner developed a system that should aid urban planners when conducting site visits by providing environmental information in the context of the physical site [WF09]. Exploring different visualizations to present localized carbon monoxide levels they assume that spheres are easier to localize than cylinders. The authors argue that virtual smoke to visualize the data has a high psychological impact. Dey et al. [DCS10] and Sandor et al. [SCDM10] explored approaches to visualize occluded POIs. Comparing different photorealistic X-ray type visualizations in outdoor environments they show that users underestimate the distance to occluded objects [DCS10]. As their results contrasts previous work that uses HMDs they assume that there are fundamental differences between depth perception in non-near eye displays and HMDs. Advanced presentation and interaction techniques for augmenting urban areas have been developed and evaluated by Alessandro et al. [ADS10]. The interface allows a smooth transition between an AR view, an egocentric panorama, and a map. They show that performance and preferences for the interfaces depend on the respective task. Therefore, it can be assumed that enabling users to smoothly switch between different presentation techniques would empower users to choose the appropriate technique.

2.5.4 Summary

It has been argued that most work in the handheld AR domain focuses on the technical foundations. Henrysson et al., for example, stated in 2007 that "there has been little research on interaction techniques for mobile phone AR, and almost no formal usability studies have been conducted." [HMB07]. The design space for handheld AR devices has been explored in different work. Investigated approaches include handheld AR binoculars [GDB07] and custom solutions based on UMPCs [VK08, WDH07]. Most work, however, is using PDAs and smartphones (e.g. [HBO05b, WPLS05, RO08, OF09]). To determine the handheld devices' position and orientation prototypes use visual markers (e.g. [WPLS05, Roh07, HBO05a, OF09]), computer vision based on natural feature (e.g. [MML$^+$11, LLLW10], or recent smartphones' inertial sensors (e.g. [KMG$^+$09, TF09]).

Results of studies that compare handheld AR with AR using HMDs suggest that both techniques can provide a similar impression [Fit93] and handheld AR can even be more useable [WDH07]. Compared to AR using HMDs, mobile devices have an inherent smaller display. Consequently, a number of studies addressed the small display size when using a smartphone for AR. It has been shown that the density of objects impact the interaction [RSS$^+$09].

Typical application scenarios that have been investigated include the interaction with paper maps [RSR$^+$07, MOP$^+$09], paper documents [LLLW10], POIs [KMG$^+$09, PCAA10] as well as using handheld AR for gaming [HBO05a, WPLS05, Roh07, OF09]. In particular, sensor-based augmentation of POIs has also gained commercial interest resulting in a number of applications for major mobile platforms. Recent sensor-based handheld AR applications include, for example, Layar[3], Wikitude[4], and Google Goggles[5] that all have been installed by millions of mobile users. Similarly, Nintendo's handheld game console 3DS can be considered as a breakthrough for using handheld AR for gaming. The Nintendo 3DS comes with preloaded handheld AR games and can be extended with additional AR games.

2.6 Discussion

This chapter provides an overview about previous work in the field of "Mobile Interaction with the Real World". After defining "Mobile Interaction with the Real World" we classify interaction techniques by distinguishing between a 1D and a 2D input space for sensing physical objects as well as differentiating techniques that provide discrete or continuous feedback. Corresponding to this classification we distinguish between the interaction techniques Touch & Hovering, Point & Shoot, Continuous Pointing, and handheld AR. We describe research that investigates the interaction techniques, discuss the used technologies, general findings, and their suitability compared to other techniques.

Use cases have been revealed where touch-based interaction is preferable compared to a mobile phone's UI [SRP08, BH10]. However, [CPV09] showed that touch-based interaction is not necessarily more efficient compared to traditional solutions and that the technical implementation can highly affect the interaction. Furthermore, results from [MBGH07] suggest that users might prefer touch-based interaction compared to Point & Shoot using visual markers. Further results from comparisons of the explicit interaction techniques touch and Point & Shoot suggest that Point & Shoot is not preferred by users [RLC$^+$06, BSR$^+$07, RLS07]. In conclusion, touch-based interaction techniques should be preferred compared to Point & Shoot if the physical objects are in the user's reach. Apparently, however, Point & Shoot can be used to interact with objects from a distance and, thus, supports use cases that cannot be addressed by touch.

It has been suggested that Point & Shoot is preferable compared to text-based search [FXL$^+$05, TMN07] but only some users prefer it compared to browsing a list [DCDH05a]. The addressed use cases might, however, not be the typical ones. Looking

[3] Layar: `http://www.layar.com` last accessed 24 November 2011

[4] Wikitude: `http://www.wikitude.com` last accessed 24 November 2011

[5] Google Goggles: `http://www.google.com/mobile/goggles` last accessed 24 November 2011

for example at advertisement posters, a typical example where Point & Shoot is already commercially used, the posters usually provides an URL that can directly be typed into a phone's browser. It therefore remains to be investigated if Point & Shoot can be more usable than simply typing an URL.

Little effort has been invested to compare the continuous interaction techniques either with the explicit ones or with traditional interaction techniques. In comparison with traditional interfaces, it has only been shown that using handheld AR and a paper map outperforms a digital map controlled by a phone's joystick. Solely the seminal work by Fitzmaurice suggest that handheld AR can provide a similar impression as AR using HMDs [Fit93] and Wither et al. suggest that using handheld devices can be even more useable than using HMDs [WDH07]. What is missing is a formal comparison of the three interaction Point & Shoot, Continuous Pointing, and handheld AR that all can address very similar, if not the same, use cases.

A number of researchers investigated the design of the labels that generates affordance for touch-based interaction and Point & Shoot (e.g. [Arn06, VTK06, YSY+09]). Basically this kind of research investigated the design of the physical UI. Since it has been argued that content-based camera-based approaches are superior [RSKH07, EAH08, HB08] explicit labels might not be reasonable. Much less work has been investigated to design the UI itself, in particular the design of handheld AR interfaces. Current handheld AR interfaces are usually the result of the designer's intuition. What is missing are guidelines for the UI design based on empirical research.

It has been highlighted that the small field of view of a phone's camera used for handheld AR is a clear limitation [TF09]. Little work has been invested to counteract the small display size if using handheld AR. The preliminary study by Pombinho et al. suggests that providing a 2D overview in addition to a handheld AR interface seems valued by users [PCAA10]. A number of other visualization techniques for virtually extending a device's screen have been developed for other interaction techniques (e.g. [ZMG+03, BR03, BCG06, BC07, GBGI08]. It remains to be examined if and how they can be applied to handheld AR.

A number of systems using Point & Shoot and handheld AR are commercially available (e.g. SnapTell, Google Goggles, and Layar). Touch-based systems have also been deployed but commercial breakthrough of touch-based interaction is clearly hampered by the lack of widely available handheld devices that support this interaction techniques. Nokia, for example, launched the world first NFC enabled mobile phone, Nokia 3220, in April 2005 [Nok06]. In the subsequent years Nokia and other manufacturers of mobile phones announced or launched NFC-enabled phones. These devices, however, did not achieve a relevant market penetration. In contrast, Point & Shoot, Continuous Pointing, and handheld AR can all be implemented using camera-based approaches. These techniques are therefore particularly viable for current devices.

3 Comparison of Techniques for Mobile Interaction with Physical Objects

For most man-made objects related digital information is available today. While on the move we can use mobile phones to search for product descriptions or other information. Entering URLs or using internet search engines to find information can be an unsatisfying experience on a mobile device. In particular because of smartphones' limited text entry capabilities. Camera-based interaction techniques that take the image of the phone's camera as an input and recognize objects have been proposed as a replacement for text-based search [DCDH05b, FXL$^+$05, TMN07]. While camera-based interaction techniques to access information about physical objects have been addressed in previous work (see Section 2), previous work is mainly of explorative and qualitative nature. It has not been shown that these interaction techniques are preferred by users or can be used more efficiently compared to established interaction techniques. In addition, different camera-based techniques can be used but it has not been studied which should be preferred when developing mobile applications.

In this chapter we analyse camera-based interaction techniques to access information about physical objects by conducting four consecutive user studies. After providing a description of the addressed interaction techniques we investigate the differences between manual text input and the camera-based interaction technique Point & Shoot in a controlled experiment. The characteristics of Point & Shoot are further analysed using a prototype to interact with printed photo books. Interaction with posters using the interaction technique Continuous Pointing is analysed in an explorative user study. The three camera-based interaction techniques Point & Shoot, Continuous Pointing, and handheld AR are finally compared in a controlled experiment. We close this chapter with a summary and an outline of the implications of our findings on the design of mobile applications that are used to access information about physical objects.

3.1 Investigated Interaction Techniques

Different approaches have been proposed to interlink physical objects with digital services using mobile phones. Camera-based approaches that use image analysis to recognize physical objects are an especially promising direction because these do not require adding additional hardware to the phone, do not depend on markers that must be attached to the object, and are already used by consumers today. Still, different camera-based interaction techniques are possible but it is not clear which one is preferred by users or can be used more efficiently. In addition, image-based object recognition is needed for this kind of interaction but different interaction techniques require different amounts of memory and processing power. In this chapter we investigate three different camera-based interaction techniques that enable to access information about physical objects. In addition, we consider the manual input of an URL using a virtual keyboard as a baseline. Therefore, the interaction flow for manual text input as well as the interaction flow using the three camera-based interaction techniques Point & Shoot, Continuous Pointing, and handheld AR is described in the following.

3.1.1 Manual Text Input

A large number of physical objects that should serve as an anchor to enable users to access additional information provide the user with an URL printed on the objects. We showed in [PHN$^+$08] for advertisement posters, for example, that with 68% the majority contain an URL. Using today's smartphones and PDAs URLs are typically entered by typing the URL with a soft keyboard.

A typical interaction flow where text entry is used to retrieve information related to an advertisement poster is depicted in Figure 3.1. After the user spots a poster he or she can decide to get more information about the advertised product or service. If so, the user opens the web browser. Using the soft keyboard the URL printed on the poster is entered. After submitting the URL the according website is retrieved.

Nowadays most PDAs and smartphones rely almost solely on touchscreens and the users' fingers or a stylus for input and output. Soft keyboards are the only well-established input technique for text that smartphones without a hardware keyboard, such as Windows Mobile PDAs, the iPhone, and most Android devices, offer. A keyboard is displayed on the screen and the user simply taps the shown keys with the finger or stylus. The user can directly see the most common and available characters. PDAs being especially popular in the 2010s relied mainly on resistive touchscreens that combined input- and output space in one screen and the user interface was optimized for stylus-based input. The usage of a stylus came almost out of fashion as it could easily be lost, it takes time to retrieve them, and their usage implies two-handed interaction. Using capacitive touchscreens that are used with the finger, however, leads to the "fat-finger-problem" [SRC05]. The output resolution of such touchscreens is much higher than the input resolution of a human thumb or finger. It is therefore difficult to select small targets with a much larger finger and the finger also occludes the target.

3.1.2 Point & Shoot

Camera-based interaction techniques that take the image of a smartphone's camera as an input and recognize objects inside the camera image can bypass the manually entering a URL or search terms. Point & Shoot is the most widely used camera-based interaction technique currently available for smartphones. The user takes a photo of an object and is provided with related content. Point & Shoot is used by a number of commercial applications such as Snaptell[1] [Ram07], Nokia Point & Find[2] [GSJ$^+$07], and Google Goggles[3].

[1] Snaptell - Visual Product Search `http://snaptell.com` last accessed 24 November 2011

[2] Nokia Point & Find `http://pointandfind.nokia.com` last accessed 24 November 2011

[3] Google Goggles `http://www.google.com/mobile/goggles` last accessed 24 November 2011

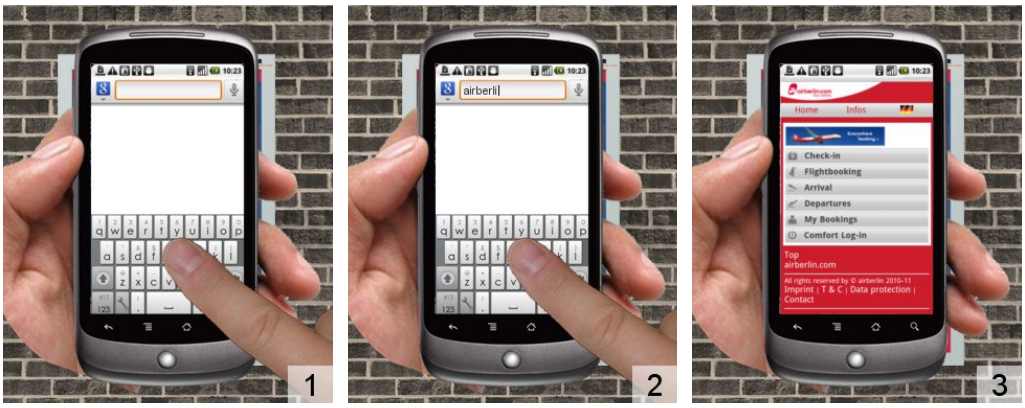


Figure 3.1: Concept of the interaction flow for manual text entry. After spotting an advertisement poster the browser is started (1) and an URL is entered (2). The user receives the according website after submitting the URL (3).

Figure 3.2 depicts the interaction flow using Point & Shoot. The user first decides to get more information about an object and starts an application that uses Point & Shoot. The camera image of the phone's camera is displayed on the screen and the user takes a photo of the object. The object recognition process is explicitly triggered by the user by pressing a button or touching the touchscreen to take the photo. Current commercial applications submit this image to a remote server that tries to match the photo with a image database. If a match is found connected metadata is transmitted back to the phone. The metadata could, for example, include a URL that leads to a website for the photographed object. Including matching the camera image with a database the transmission of the data between phone and server takes around three seconds using commercial application.

Implementations of server-based Point & Shoot are based on robust image matching algorithms. The algorithm must be invariant or robust against scaling, rotation, partial occlusion, perspective transformation, blur and other noise. Different image features [TM08], such as SIFT [Low99, Low04], SURF [BTVG06, BETVG08], and MSER [MCUP04] have been proposed for this purpose. Combining these low-level image features with an additional heuristic, such as using a vocabulary tree for image recognition developed by David Nistér and Stewénius [NS06], enables to recognize objects with logarithmic complexity. E.g. [NS06] shows that matching an image with a database containing 1,000,000 images took about one second on a powerful desktop computer in 2006.

3.1.3 Continous Pointing

Point & Shoot can be extended to Continuous Pointing by analysing the stream of images provided by the phone's camera and not just a discrete photo. Figure 3.3 depicts the interaction flow using Continuous Pointing. As with Point & Shoot the camera image is

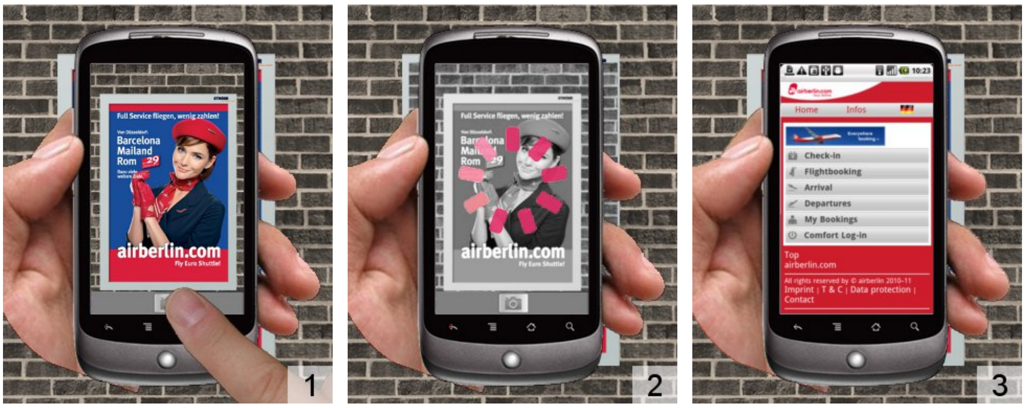


Figure 3.2: The interaction flow using Point & Shoot. After spotting an advertisement poster and starting the application the user aims at the poster (1). The user takes a photo that is analysed (2) and receives the according website if the poster has been recognized (3).

shown on the phone's display. The stream of images delivered by the phone's camera is continuously analysed using object recognition algorithms. If the user aims with the phone at an object and the object appears in the camera image the object is automatically recognized. No explicit interaction by the user is required. If an object is identified according metadata is displayed on the screen.

Assuming a perfect object recognition algorithm the information about an object is shown by the phone as soon as an object emerge in front of the camera. The user receives direct feedback and does not have to explicitly trigger the recognition. Continuous Pointing can be implemented by either transmitting the stream of camera images to a server or by analysing the images directly on the phone. Transmitting the images to a server requires a high network bandwidth. E.g. a stream of uncompressed black and white images with 240x320 pixels and 25fps requires transmitting 1.83mb of image data per second. This is more transmitted data per second than Point & Shoot requires for selecting one object. Assuming that a user selects one object every minute Continuous Pointing would require 30 times more bandwidth. As, for example, shown in [HSB09] the object recognition can also be implemented on a mobile device. While today's smartphones' processing power is sufficient the available memory limits the number of recognizable objects to a few hundreds.

3.1.4 Handheld Augmented Reality

If the camera image is continuously scanned for objects it is only a small step to handheld AR. As with Continuous Pointing the objects inside the camera image are automatically recognized. By determine an object's position inside the camera image the object can be augmented. Figure 3.4 exemplary shows the interaction flow using a handheld AR interface that just highlights the object. Using handheld AR the user not only receives

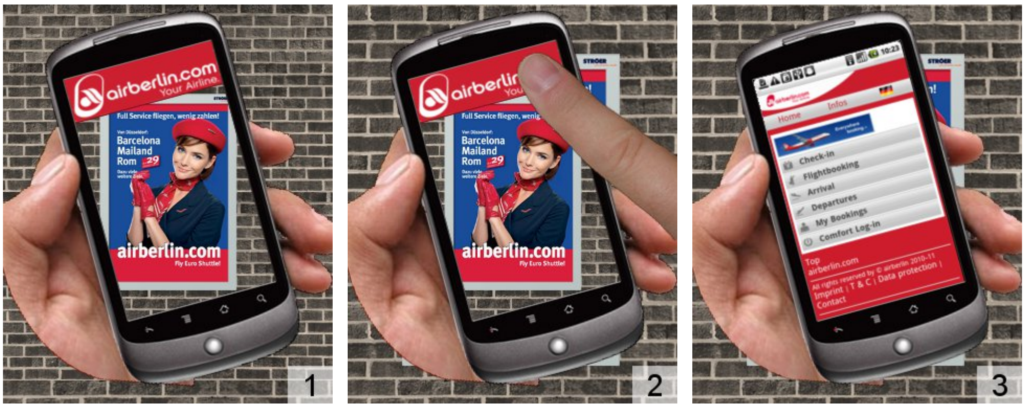


Figure 3.3: The interaction flow using Continuous Pointing. After spotting an advertisement poster and starting the application the user aims at the poster. The camera image is constantly analysed and a notification is displayed on the phone's screen as soon as an object is recognized (1). The user can select the notification (2) and receives the according website (3).

feedback that an object inside the camera image is recognized but the object can also be highlighted. This avoids ambiguity if multiple potential physical objects are simultaneously visible in the camera image. Presenting information aligned to the objects' position enables to present information about multiple objects simultaneously.

An implementation of handheld AR can be based on similar algorithms as Continuous Pointing. In addition, it is necessary to determine the pose of recognized object. This involves determine a coarse pose that is iteratively refined. As most mobile phones are equipped with a slow floating-point unit or no floating-point unit at all the pose estimation algorithms must rely on fixed-point numeric [WRM$^+$08]. Therefore, it is difficult to find a good balance between determine a precise pose and speed. In general, determine the objects' pose requires additional processing time. For the algorithm proposed by Wagner et al. [WRM$^+$08], for example, outlier removal and pose refinement, both required to determine an objects pose, took 11% of the overall time per image. 11%, however, are only sufficient if objects are constantly tracked with a high frame rate.

3.2 Comparison of Point & Shoot and Manual Input

The camera-based interaction techniques to access information connected to physical objects described above are only valuable if they provide a benefit for the user compared to other interaction techniques. Therefore, we started the analysis of the camera-based techniques by comparing them with a baseline condition. As described in Section 3.1 different camera-based interaction techniques are available. As we assumed that Continuous Pointing and handheld AR are superior compared to Point & Shoot we conducted a controlled experiment that compares Point & Shoot with manual text input that serves as a baseline.

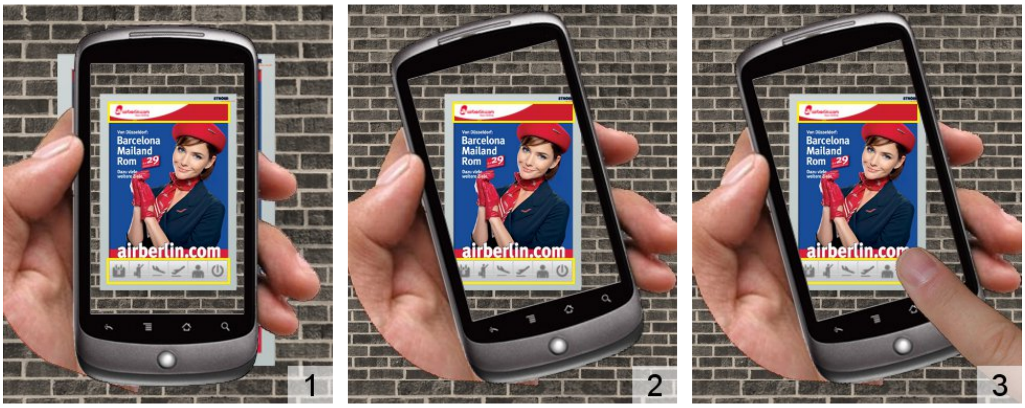


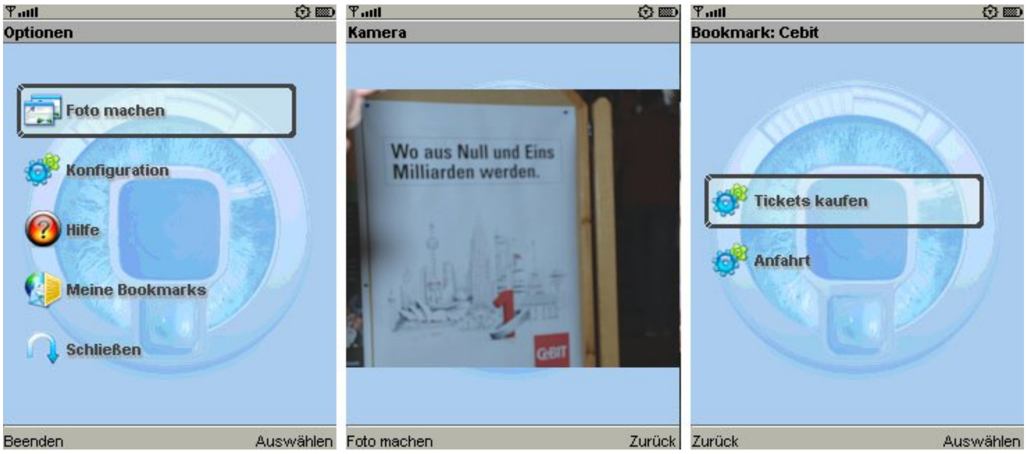
Figure 3.4: The interaction flow using handheld AR. After spotting an advertisement poster and starting the application the user aims at the poster while the camera image is constantly analysed. The image is augmented with information and buttons (1). Information and buttons are directly presented aligned to the object (2). Information directly augments the object and buttons on the object can be selected to activate functions or retrieve further information (3).

Based on our analysis of advertisement posters [PHN$^+$08] we found that most of them contain an URL. We therefore assume that advertising companies have a strong interest in leading potential customers to their website. This makes advertisement posters a promising type of physical object and hence we selected advertisement posters as the type of physical objects for the study. In the experiment, users had to access information related to advertisement posters using both Point & Shoot and manual text input. Our assumption was that the camera-based technique is faster and easier to use than text-based interaction techniques provided by today's devices.

3.2.1 Developed Prototype

In order to conduct the study we implemented a system that enables to access services related to posters using content-based image analysis (see also [PHN$^+$08]). The system consists of three components: Posters that we consider as a specific type of real world objects, a mobile camera phone that provides the user interface, and a server application that stores descriptions and related services for posters.

The mobile phone centralizes the interaction between the user and the system. The user creates a photo of a poster with the application. This photo is transmitted to a server application that compares it with images of posters stored in a media repository. For this comparison Scale Invariant Feature Transform (SIFT) keypoints [Low04] keypoints are extracted from the photo as well as from digital images of the posters. The SIFT keypoints are compared pair wise and the poster with the highest number of matching keypoints is considered as the photographed poster. For each poster, descriptions of related services consisting of a category (e.g. ticketing service, background information,



(a) (b) (c)

Figure 3.5: Screenshots of the mobile application for using Point & Shoot with advertisement posters. (a) Application's main menu, (b) view to select a photo by taking an image, and (c) links to two services are provided.

and opening time), a name, and an according URL are stored in conjunction with the respective poster. The description of the poster with the highest number of matches is transmitted back to the mobile phone. On the phone the services are presented to the user. By selecting a service the user is guided to an according web page. Three screenshots of the mobile application are shown in Figure 3.5. For the study the prototyped was only trained with three reference posters. Therefore, it was very robust and we did not experienced wrong matches throughout pilot testing and the study.

3.2.2 Method of the study

In the controlled experiment participants executed a single task to compare Point & Shoot with entering an URL using a soft keyboard. A within-subject design with one independent variable and two conditions is used for the tasks.

3.2.2.1 Design

The interaction technique is the experiment's independent variable. Using repeated measures we counterbalanced the order of the conditions to reduce sequence effects. Dependent variables are the task completion time and participants' subjective ratings. We measured the task completion time by stopping the time participants need to answer a question related to the content promoted by a poster. In addition, we asked them to rate the interaction techniques' ease of use on a five point Likert scale using a questionnaire. Furthermore, participants provided additional information and subjective impressions also using questionnaires.

3.2.2.2 Participants

46 people (23 female) participated in the experiment. Most participants were students of the University of Oldenburg. They were 20-50 years old (M=25.16, SD=5.67). 93.3% of the participants owned a mobile phone and 92.9% of these phones are equipped with a camera. 57.8% of the participants have at least once installed an application on their mobile phone.

3.2.2.3 Apparatus and Materials

The apparatus consists of three advertisement posters, two Nokia N95 mobile phones, two Glofiish PDAs, a laptop, and a WLAN access point. At the time the study has been conducted, the Nokia N95 had one of the best cameras integrated in a mobile phone. The Glofiish PDA had a large touch screen that made text entry much easier compared to using the Nokia N95. Thereby, we ensured that the performance using the soft keyboard condition is not negatively biased by the poor text entry support offered by the Nokia N95.

Each of the three different advertisement posters contained a short clearly visible URL (e.g. cebit.com). We intentionally selected very short URLs that can easily be typed to create optimal conditions for entering them using a soft keyboard. Content optimized for mobile Internet browsers was prepared for all posters.

The image matching server was installed on the laptop. The laptop also served as a web server to provide web pages related to the posters. An additional DNS server on the laptop allowed entering standard URLs such as cebit.com to access the prepared web pages. The WLAN access point was connected to the laptop to ensure reliable wireless network access for the mobile devices.

The client application was installed on the Nokia N95s (see Figure 3.6 b). Participants took photos of the posters using the integrated high resolution camera. According to the server's response the Nokia N95 displays a list containing five web-links to information related to the respective poster. For our control condition we used a Glofiish X500+ PDA (see Figure 3.6 a) with a soft keyboard running Windows Mobile 2005. Entering a URL printed on a poster in the integrated web-browser the web-server returned an according web-page containing links to the same pages used for the mobile phone.

3.2.2.4 Procedure

We set up the evaluation booth at the student cafeteria of the University of Oldenburg (see figure 3.7). The study was conducted on a Monday from 11.00 to 15.00 during the term. People were randomly asked to participate in the study. After a person had agreed to participate in the evaluation, the experimenter made the participants familiar with the posters and outlined that he or she had to answer questions regarding the events advertised by the posters.

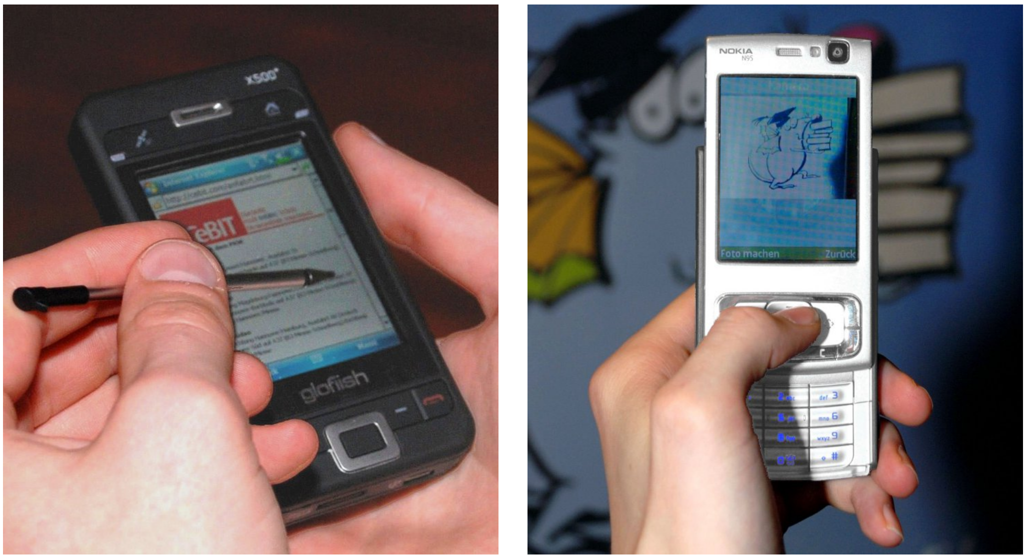


Figure 3.6: The two systems to compare manual text input with Point & Shoot: Participant enters an URL on a PDA (left) and participant shoots a photo of a poster (right).

Each interaction technique was introduced to the participants before starting the experiment. Both devices were handed out with the relevant application running. After the interaction technique for the first run was chosen the participants were asked to answer a question regarding one of the promoted events (e.g. "What are the opening hours of the CeBIT?"). A stop watch was started right after the device was handed out and the question was asked. When the participant gave the answer, the stop watch was stopped and the time was noted down. Then the participant switched to the other interaction technique and the procedure was repeated, with a different question relating to another event. Afterwards, the participants were handed out the questionnaires, where they could rate both interaction techniques. The Glofiish's browser cache was deleted after each participant, so the next participant had to re-enter the URLs in full length.

3.2.3 Results and Discussion

The average time to answer the given question was 52.81s (SD=20.88) using Point & Shoot and 64.88s (SD=25.08) using the soft keyboard (see Figure 3.8). The interaction technique had a significant but small effect on the response time. Participants needed less time to complete the task using Point & Shoot compared to using the soft keyboard (p<.01, r=.16). The interaction techniques' ease of use was rated on a five point Likert scale, where 0 meant very easy and 4 meant very difficult. The average rating of Point & Shoot was 0.94 (SD=1.00, Mdn=1) and the average rating using the soft keyboard was 1.62 (SD=1.02, Mdn=1.5) (see Figure 3.8). A Wilcoxon signed-rank test revealed a significant large effect on the perceived ease of use (p<.001, r=.61). Participants found Point & Shoot easier to use than using the soft keyboard. Asking participants if they

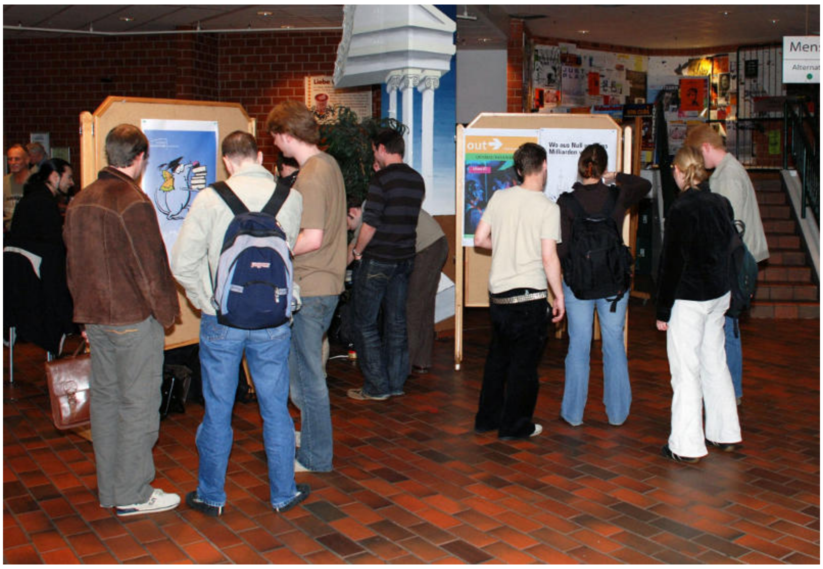


Figure 3.7: Evaluation booth for the study that compares Point & Shoot with manual text input in the student cafeteria of the University of Oldenburg.

would be willing to pay for the demonstrated service we found that only 4.4% were willing to pay money for the service, 37.8% might be willing to pay, and 57.8% negate to pay.

The main result of the evaluation is that accessing information connected to advertisement posters using Point & Shoot is significantly faster and is considered as significantly easier to use by participants. However, the effect on the response time was statistically small. We argue that this finding is still substantial, since the non-clinical setup increases the external validity but also increases the variance. For example, participants were not patient enough for a longer training session, so all of them were untrained. Participants also spent part of the time browsing the web-pages themselves rather than performing the interaction for accessing them. We expect that a lab study just comparing both interaction techniques would reveal a larger effect size. Furthermore, the procedure, in particular that URLs were printed on the posters, provided ideal conditions for the text-based condition. The time to type a URL depends on the number of characters that must be typed. We intentionally selected very short URLs (e.g. cebit.com) that can easily be typed. Specific use cases will require longer URLs and therefore further increase the time that is required to type them. In addition, recent devices offer a higher photo quality and are able to take photos faster. This would even further increase the advantage for Point & Shoot.

Our results are consistent with the findings of Davies et al. [DCDH05a]. They conducted a field trial to compare Point & Shoot-based interaction with a dialog system [DCME01] in the context of mobile tour guides. They conclude that users appear happy to use Point & Shoot even when "this is a more complex, lengthy and error-prone pro-

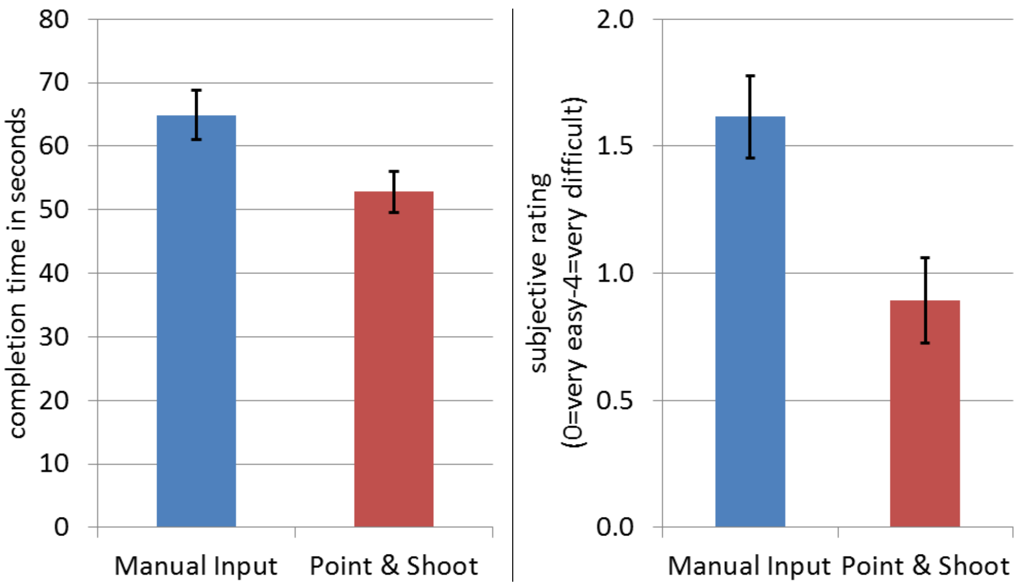


Figure 3.8: Task completion time (left) and participants' subjective rating (right) using a soft keyboard and Point & Shoot (error bars show standard error).

cess than traditional solutions". Davies et al. also conclude that "developers should not be concerned about user acceptance of digital image recognition techniques for object identification". Thus, we were able to show that Point & Shoot provides a clear benefit for users compared to entering a URL using a soft keyboard to access information related to advertisement posters and Davies et al. came to similar conclusions in the context of mobile tour guides. We created near optimal conditions for the manual text entry condition and, therefore, assume that our findings can be generalized beyond a particular type of physical object. In particular if considering that other camera-based interaction techniques might provide an even larger benefit compared to Point & Shoot.

3.3 Point & Shoot Interaction for Printed Photo books

In order to further investigate the characteristics of Point & Shoot we conducted a user study for another type of physical object. During this study users selected photos in a printed photo book using Point & Shoot. Besides collecting feedback from potential users we were also interested in determining the accuracy of the image matching when the system is used by ordinary persons. We therefore analysed the percentage of correctly recognized photos and measured participants' impressions using questionnaires.

3.3.1 Developed Prototype

We developed a system we call "Bookmarkr" that provides a link from physical photos to their digital counterpart. Thereby, the system can provide metadata for physical

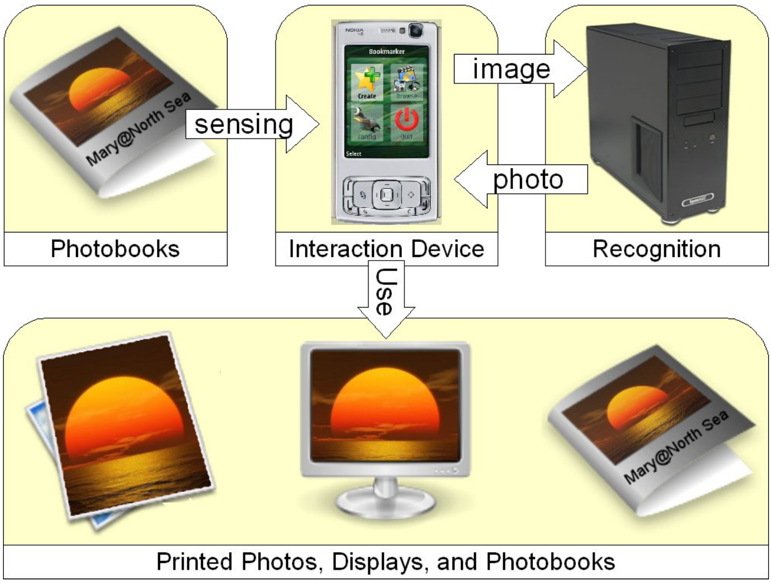
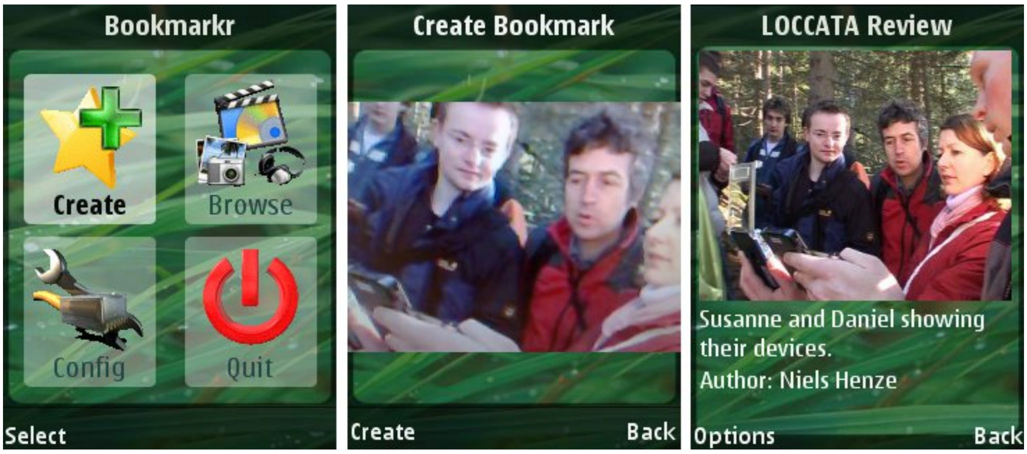


Figure 3.9: Architecture of the Bookmarkr prototype for using Point & Shoot to interact with printed photo books.

photos that might be available in photo sharing communities such as flickr.com and photobucket. In particular, users can collect and takeaway the digital version of a printed photo easily. The system does not require major changes of the design or cost of individual photo books. With the developed process (see Figure 3.9) users can access printed photos' digital counterpart using a mobile phone application. Using Point & Shoot the user creates an image of the printed photo book using the phone's camera. The image is sent to a photo book server which accesses a photo repository containing the digital versions of the photos. This server could, for example, be located at the user's photofinisher company that prints the photos to paper and thus has access to the digital photos. The photo book server retrieves the photographed picture by matching the image taken by the mobile phone with the digital photo. The retrieved digital photo is sent back to the mobile phone.

Designing the system requires taking into account the enormous amount of digital photos on photo sharing websites today and the amount of photo books produced each year. It can be considered as impossible to retrieve the correct photo in a reasonable amount of time and with sufficient precision for every photo ever created. We therefore envisage a two stage process. The user first selects a photo book before he or she can select photos from the book. We envisage the same interaction technique for selecting a photo book as for selecting individual photos. Photo books are selected by taking an image of its cover page[4]. Thus, images taken by the user are compared with all

[4] Considering the amount of printed photo books it might be necessary to use other techniques to reduce the search-space. This can be achieved by taking an image of the individual barcode on the back side of each photo book instead of taking an image of the cover page.



Susanne and Daniel showing their devices.
Author: Niels Henze

(a) (b) (c)

Figure 3.10: The Bookmarkr photo sharing application running on a Nokia N95. (a) Application's main menu, (b) view to select a photo by taking an image, and (c) digital photo is displayed in conjunction with metadata.

previously selected photo books and all cover pages (or barcodes on the back side of each photo book).

Retrieving a digital photo based on an image taken by a mobile phone's camera is the core of the overall process. We extended the standard image matching algorithm SIFT [Low99, Low04] to implement a responsive system. The algorithm takes the query image and extracts SIFT features. These features must be compared with the SIFT features extracted from all digital photos. The desired outcome of the matching is the most similar photo and not a collection of similar photos. For the image matching algorithm it is therefore sufficient to determine the matching SIFT features of the whole photo collection and it is not necessary to determine the matching feature for of each photo. On this basis, we extended the standard algorithm used in the previous study (see [HB08] for more details). All SIFT features for one photo collection are stored in a single kD-tree. This decreases the matching time dramatically compared to using a kD-tree for each individual photo used in previous work. It must be taken into account that using a best-bin-first heuristic, that is also part of the standard SIFT process, decreases the robustness of the process. The robustness is further decreased by larger kD-trees resulting from storing keypoints from multiple photos in a single tree. However, as we will show in Section 3.3.3 and [HB08] this is, at least for our application, not relevant. The outcome of the pre-processing is one kD-tree for each photo collection each containing the SIFT keypoints of the photo collection. The kD-trees are stored alongside the digital photos in the media repository.

We implemented the server application including the matching process described above. We use a combination of Microsoft's C# and C using Rob Hess's SIFT implementation [Hes08] that extends Intel's Computer Vision Library OpenCV [Int08]. The server application runs on Microsoft Windows XP. All parts of the system that are not performance critical, for example the network communication, are implemented in C#. The performance critical parts, in particular the SIFT algorithm and the keypoint matching, are implemented with C using Microsoft's C compiler provided by Visual Studio 2005. The mobile client application is implemented using Java ME (MIDP 2.0, CLDC 1.1). Communication between the mobile application and the photo book server is established via a TCP/IP connection. The retrieved digital photos can be sent to other devices, stored in the device's file system, and deleted. Figure 3.10 shows three screenshot of the mobile application. Upon starting the application the main menu is displayed. By selecting the "Create" item the camera's view is displayed on the screen. After shooting an image of a photo to select it the digital photo is displayed in conjunction with metadata. Using this view's option menu the photo can be transferred to others. Users can also browse through a list of photos they have collected.

3.3.2 Method of the study

Using the Bookmarkr prototype described in the previous section we performed two evaluations. Synthetic experiments that are described in [HB08] were conducted to determine the performance of the developed system. We conducted the user study described in the following to analyse the performance of the developed algorithm under realistic conditions and investigate the suitability of the system for the targeted audience.

3.3.2.1 Design

In the user study the system was used to determine the accuracy of digital photo retrieval when used by ordinary persons. The user study focuses on using a Point & Shoot interaction technique to retrieve photos from a photo book. Thus, participants did not have to select the photo book first. In addition to objective data we also collected subjective feedback from potential users. We analysed the percentage of correctly recognized photos and measured participants' impressions using questionnaires.

3.3.2.2 Participants

Our group of ten participants was diverse with regard to gender, age, and mobile application experience. The participant population consisted of an equal split of males and females. Two participants were between 20 and 30 years old, four participants were between 30 and 40 years, two between 50 and 60 years, and two were between 60 and 70 years. Two participants had no experience with digital photography and did not own a mobile phone. All other participants had experience with digital photography. Four of them stated that they share a mobile phone with their partner and rarely use it. The others use a mobile phone on a regular basis.

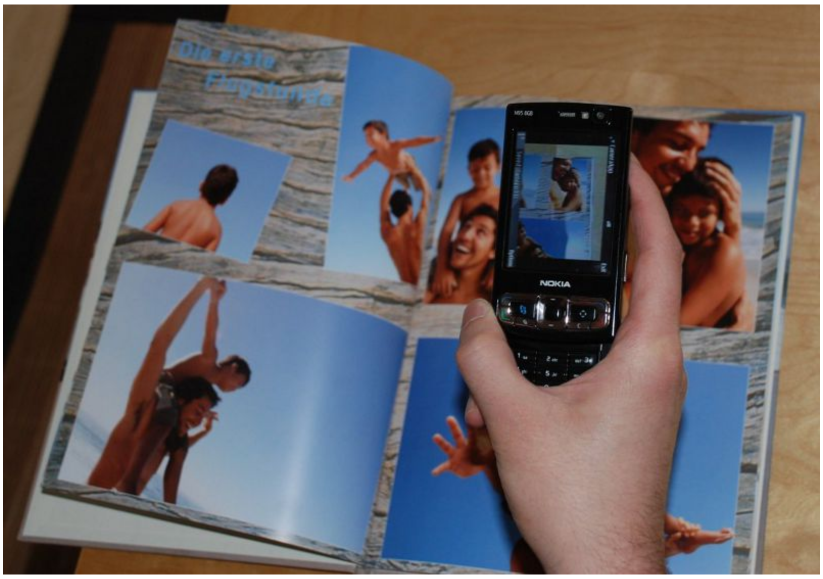


Figure 3.11: User taking an image of a part of a photo book's page using a Nokia N95.

3.3.2.3 Apparatus and Material

In order to establish the link between the printed photo book and digital the system described in 3.3.1 was used. An Apple MacBook running Microsoft Windows XP equipped with a 2 GHz mobile Intel Core2Duo processor and 1 GB memory was used to run the server application.

Participants used a Nokia N95 8GB, running the mobile application to take images of the photo book (see Figure 3.11). The images taken with the Nokia N95 had a resolution of 960x720 pixels. The communication between the Nokia N95 and the server application was established using a local Wi-Fi connection.

A representative photo book was designed for the study. The photo book contained 46 photos. The book has been printed by a photofinisher company. The photos of the photo books were pre-processed with a resolution of 1024x768 pixels.

3.3.2.4 Procedure

The experiment was conducted in a living room styled setting at different daytimes and without electric light. The room had one window. Figure 3.12 shows the setting of the evaluation. Each participant was welcomed and asked to take a seat on a sofa. The aim of the system and the evaluation procedure was explained to the participants. They were not asked to pay special attention while taking the images. Afterwards the participants were provided with the photo book and a Nokia N95 mobile phone running the mobile application. The assistant asked the participants to select at least ten photos using the application.

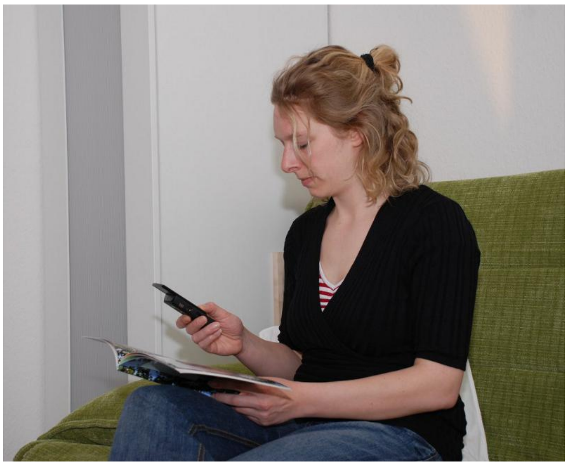


Figure 3.12: Participant during the evaluation using a Nokia N95 to select photos from a photo book.

After selecting the photos participants were asked to complete a questionnaire. Besides demographic questions the questionnaire consisted of multiple choice questions, ratings on a five point Likert scale ranging from zero (never/difficult) to four (often/easy), and an open question for additional comments.

3.3.3 Results and Discussion

Each participant took twelve images on average resulting in 121 images. For 117 taken images the system returned the correct result (96.7%). For the remaining four images (3.3%) the system returned no result. Two participants signalized that they would like to stress test the system's capabilities (e.g. "Let's see if this also works") before taking an image that could not be recognized. Figure 3.13 shows four examples of images taken by the participants. On the left are two examples that were correctly recognized and on the right are two examples that could not be recognized. Another photo taken during the evaluation that was correctly recognized is shown in Figure 3.14. In general the vast majority of the taken images are very blurred. It could be observed that most images were close-ups consisting of only a fraction of the respective printed photo.

Two participants mentioned that they are afraid of using the system after the system was explained. They stated that they are not used to handling mobile phones. For these two and an additional participant it was difficult to handle the mobile phone with one hand. They repeatedly held their fingers in front of the phone's camera while trying to take an image. Apart from holding the phone and pressing its keys no participant had difficulties in using the system. For one participant, however, the introduction to the system was not sufficient. After taking four images she realized that the photos shown on the phone's display are not the images she took with the phone's camera but their original digital counterpart.

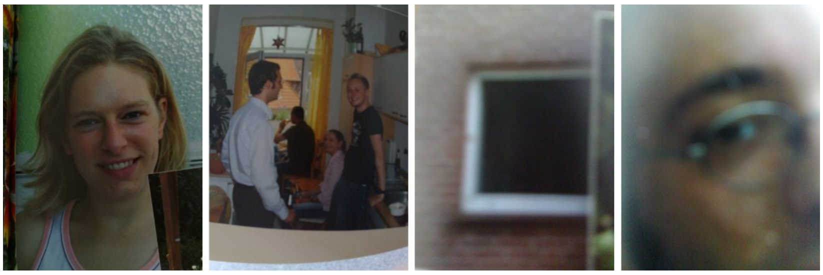


Figure 3.13: Examples of images taken with a Nokia N95 during the user study that investigates Point & Shoot for printed photo books.

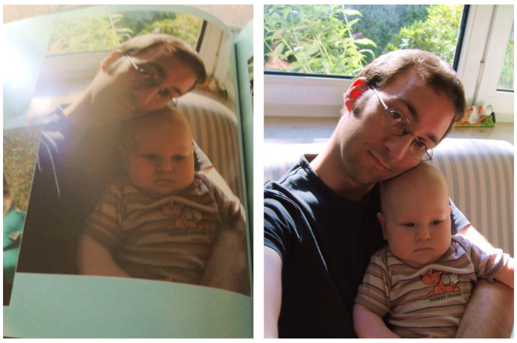


Figure 3.14: Image of a photo in a photo book taken with a mobile phone's camera (left) and the digital photo used to produce the photo book (right).

The questionnaires revealed the following insights. All but two participants rated the system as easy or very easy to use (M = 3.2, SD = 0.79). Most participants could envision using the system (M = 2.5, SD = 1.08). They would like to use the photos received on the mobile phone to order printed photos (8 times), send them via e-mail (5 times) or to their desktop computer (2 times), and view them on the mobile phone (2 times). Tools participants use to show their photos at home are diverse. Most often mentioned tools are printed photos (9 times), notebooks (7 times), digital cameras (7 times), and desktop computers (3 times). Participants stated that they would like to receive photos from others either occasionally or often (M = 2.3, SD = 0.82). They are occasionally promised to receive photos they never obtained (M = 2, SD = 1.15). Most participants envision letting others use the system to annotate their photos (M = 2.9, SD = 0.99).

The system returned the correct result for 96.7% out of 121 images. Subtracting the two images that were only shot to stress test the system's capabilities the system correctly returned 98.3% out of 119 images. The results are comparable (see Figure 3.15) with the performance we found by performing a systematic benchmark [HB08] to test the error rates for images with different resolutions. The results can be considered a promising result in particular regarding the blurriness of the taken images. One cause for the blurriness is the auto-focus mechanism of the Nokia N95 that is not able to focus on very close objects. At the time we conducted the study we assumed that the auto-focus abilities of mobile phones will improve in the future. In fact, the quality of mobile phone cameras clearly improved and current mid-range devices provide clearly superior

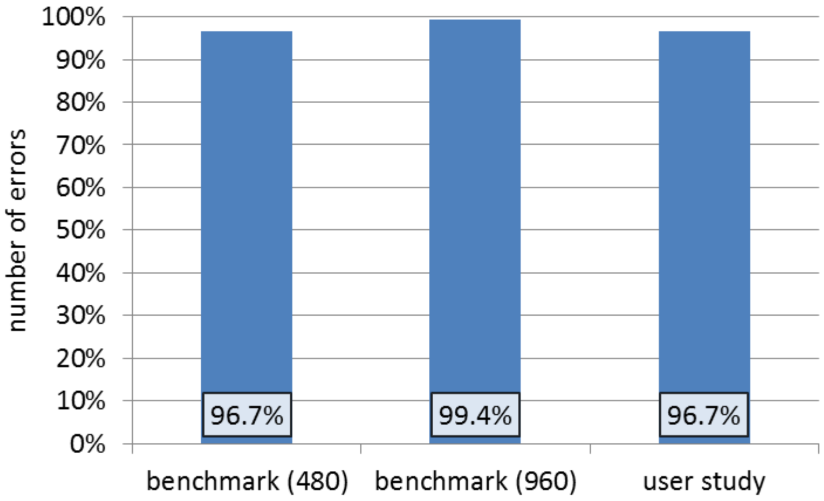


*Figure 3.15: Comparing the system's performance in the systematic benchmark we described in [HB08] and the results of the user study. The left two bars visualize the error rate for photo collections with 480x640 pixels and 960*1280 pixels.*

auto-focus abilities. Specific devices, such as Samsung's *Galaxy Camera* and Nokia's *808 PureView*, even offer a photo quality on the level provided by current compact cameras.

It could not be observed that different light conditions affect the robustness of the system. However, the influence of light condition was not explicitly tested but user's need proper light condition to view the photo book itself anyway. We assume that the light conditions that are sufficient for the users are covered by the system. SIFT keypoints can be matched robustly under different light conditions and using the phone's integrated flash light would also be possible for photo books that are not printed on high-gloss paper.

A remarkable aspect is that all participants even the elderly that do not own a mobile phone nor use a computer were able to understand and use the system. However, in particular older persons might not see the system's benefit because many of them believe to have no use for digital photos in general.

3.4 Evaluation of Continuous Pointing Interaction

In order to prepare a comparison of the camera-based interaction techniques we conducted an explorative study to investigate the implications that accompany a Continuous Pointing interaction with physical objects. In addition, we were interested in participants' reaction if the system fails to recognize parts of an object. Again posters are used for the study but in contrast to the study described in Section 3.2 one poster contained a number of regions that each can provide different content.

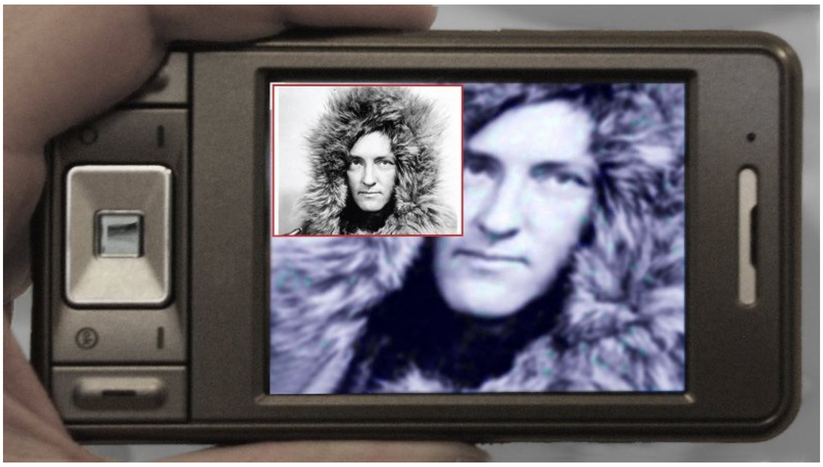


Figure 3.16: ASUS smartphone running the prototype developed for the Continuous Pointing interaction.

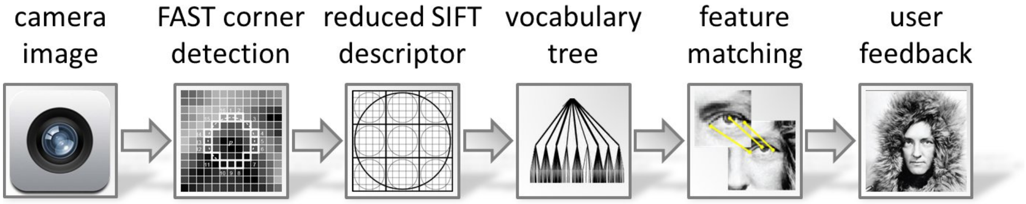


Figure 3.17: Overview of the recognition pipeline that enables to recognize a number of objects directly on the handheld device.

3.4.1 Developed Prototype

Widely used object recognition approaches such as SIFT are too expensive for today's phones in terms of processing power to analyse camera images with a high frame rate. Wagner et al. describe a simplified SIFT algorithm to estimate the 3D pose of a 2D object [WRM$^+$08]. Their approach is capable to process camera frames with a size of 320x240 pixels at a rate up to 20Hz. However, only results from recognizing a single image are reported and it was not analysed how the algorithm performs with an increasing number of objects.

We extended the approach described by Wagner et al. [WRM$^+$08] using a scalable vocabulary tree [NS06]. Our recognition pipeline is outlined in Figure 3.17. For the keypoint detection and feature description we build upon the work by Wagner et al. using a FAST corner detector [RD06] for detecting keypoint and simplified SIFT descriptors.

Wagner et al. employ a "Spill Forest" (a combination of a number of Spill Trees [LMGY04]) to match features extracted from the camera image with features from all scale steps of the reference image. Since our aim is to recognize a number of images we employ a different approach. Vocabulary trees [NS06] are able to reduce the problem to find the matching object by multiple magnitudes. Nister and Stewenius trained a vocabulary tree which reduced the problem to find an image out of two million to a problem to find an image out of a hundred candidates. The vocabulary tree described by Nister and Stewenius has a size of hundreds of megabytes and must be stored in RAM for performance reasons. We downsize the tree by reducing its level to five instead of six and a branching factor of eight instead of ten. In addition, our descriptor has only 36 entries instead of 128. Through this our empty tree needs only two megabyte. We trained our vocabulary tree with 10000 mainly high quality photos. Reference images are inserted into the vocabulary tree by extracting the features from each of the image's scale steps. Each scale step is then treated individually and inserted into the tree. By treating the scale steps individually we obtain not only object candidates from the vocabulary tree but scale step candidates. During the online-phase three scale-step candidates are retrieved from the vocabulary tree.

Since we do not aim at fine grained pose estimation no sophisticated feature matching is necessary. Thus, we rely on simple brute-force matching to compare the 100 features from the camera image with the 300 features from each of the three scale-step candidates using the sum of squared difference. To further reject potential outliers we compute a difference of orientations histogram for each candidate's matches. If this histogram shows a consistent rotation and the respective candidate's number of matches is above a certain threshold in two consecutive camera images the according image is considered as a match.

The algorithm was implemented for Windows Mobile 6 devices using *C* and *XScale* assembler. We tested the speed using an ASUS P535 smartphone (see [HSB09] for more details). The overall time to process an image from the camera takes 100ms. That means that object can be recognized with 10Hz In order to conduct the user study we developed a simple prototype shown in Figure 3.16. The prototype displays the camera image in full screen. The camera image is constantly delivered into the recognition algorithm described above. If an image is recognized a small thumbnail of the recognized image overlays the camera image. The user can get details about the object by clicking the thumbnail with her finger.

3.4.2 Method of the Study

Using the described prototype we conducted an explorative study to investigate the Continuous Pointing interaction technique. The aim of the study was to find initial indicators for designing a system using Continuous Pointing and according challenges. We monitored the users' behaviour. In particular we investigated the user's reaction if the user selects a poster's annotated regions and if the object recognition fails.

3.4.2.1 Design

The evaluation consisted of two tasks. During the first task we monitored the users' behaviour and collected quantitative and qualitative feedback. The second task is a controlled experiment with one independent variable. In the experimental condition the poster contains unrecognizable regions while in the control condition all regions can be recognized. We used repeated measures for the second task and counterbalanced the conditions. We asked the participants to fill a NASA TLX after executing the second task with each condition.

3.4.2.2 Participants and Apparatus

As we aimed at collecting early feedback only a small number of participants that share a similar background took part in the study. Six male subjects participated in the study. All participants are computer scientists and are experienced smartphone users. They were between 25 and 35 years old.

Three different posters were prepared for the study. In the first task the 45x55 cm large poster shown in Figure 3.18 was used. The poster sketches a street setting and contains seven interactive regions. If a participant selects one of these regions the phone displays advises about how to behave in the respective traffic situation.

Two very similar posters were used for the second task. Each poster contained 24 clearly identifiable interactive regions. Figure 3.18 shows a cut-out of one of the posters. For some of these regions the thumbnail that was displayed, when a region was recognized, contained a question mark. If the participant clicked the thumbnail the phone displayed either a happy green or a sad red emoticon but each poster contained only one happy emoticon.

3.4.2.3 Procedure

The study was conducted in an office room. During the first task the used poster lay flat on a table. The participants were asked to find all interactive areas without knowing its number or position. It was up to the respective participant to decide when to end this task.

For the second task the two posters were hanged on the wall. Participants' task was to find the interactive region that is connected with a happy emoticon among all interactive regions. For the first poster all regions are recognizable while for the second poster the recognition of three regions was deactivated. Participants had to execute the task using both posters. After finishing the task with one poster participants were asked to fill a NASA TLX questionnaire. The order of the posters was counterbalanced to reduce sequence effects.

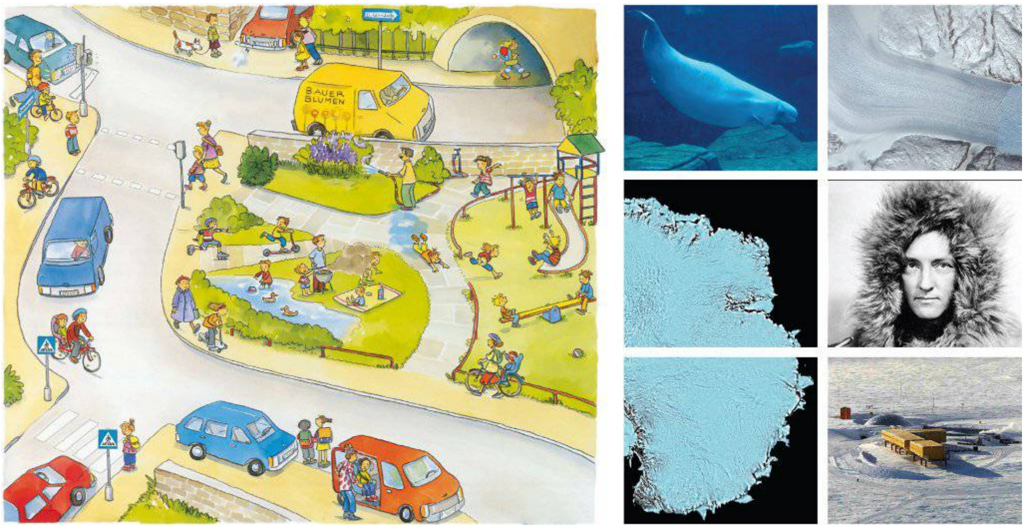


Figure 3.18: The poster used in the first task of the evaluation of Continuous Pointing (left) and an extract of the poster used in the second task (right).

3.4.3 Results and Discussion

In the first task the participants found between three and six of the seven interactive regions (M=4.66, SD=1.21). No participant was able to find the region located in the upper right of the poster. All but one participant started by systematically scanning the poster in zigzag. After scanning the whole poster once some started to scan specific regions of the poster. All but one participant permanently aligned the phone with the orientation of the poster. Three participants used the phone with one hand and in an upright posture while the others showed an inconsistent behaviour. Three participants held the phone in an almost constant height. Two participants mentioned that additional hints to surrounding interactive regions would be helpful and one participant said that it is difficult to remember the parts of the poster that were already scanned.

All participants managed to successfully complete the second task. However, probably due to the small number of participants the NASA TLX showed no significant difference between the two conditions (p=0.4, r=0.08). On average the NASA TLX score for the poster with only recognizable regions was 36.66 (SD=16.13) and the NASA TLX score for the poster with three unrecognizable regions was 39.00 (SD=13.56) (see Figure 3.19). Some participants rushed through these tasks and two did not even notice the three deactivated regions. The longest time a participant tried to select one of these regions was around 20s. All but two participants permanently aligned the phone with the orientation of the poster. One participant rotated the phone by 90° and one participant did not show a consistent behaviour. All but one participant focused most of the time on one region after the other so that the respective region approximately filled the phone's screen.

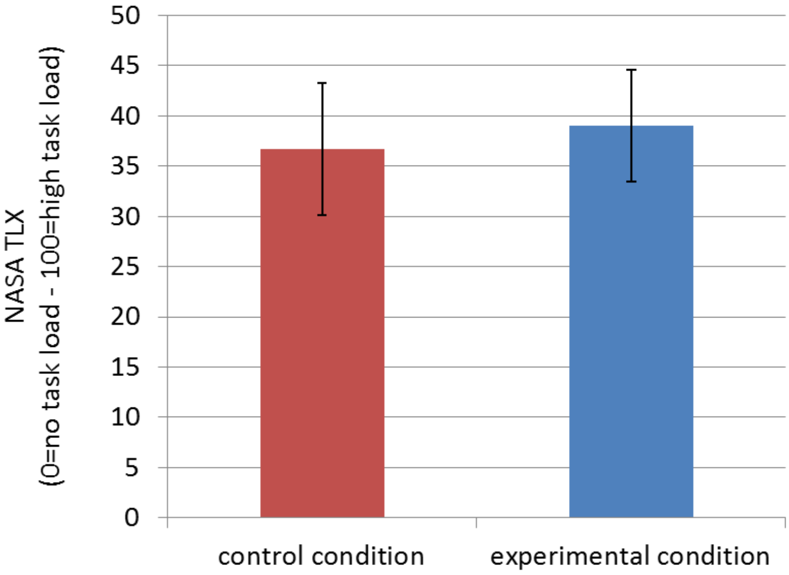


Figure 3.19: NASA TLX scores for comparing Continuous Pointing using a poster with 24 recognizable regions and a poster with unrecognizable region. Low values mean no task load and high values mean high task load (error bars show standard error).

Because of the used methodology, the small number of participants, and the participants' background the study can obviously not be generalised. However, the estimated effect size (r=0.08) suggest a very small effect. Even using a larger sample we would probably not have been able to show a significant effect. The small effect size found in a highly controlled study suggests that the difference might in fact be negligible. The results further indicate that users intentionally align the phone with the object. This is consistent with the observation we made in earlier work. It could imply that the recognition pipeline can be simplified by removing orientation invariance in tasks such as ours. Unsurprisingly the participants had problems to find all interactive regions if these are not clearly distinguishable. When marking interactive objects is not feasible additional hints displayed by the phone could ease finding nearby objects.

3.5 Empirical Comparison of Camera-Based Interaction Techniques

Different camera-based approaches have been proposed to interlink physical objects with digital services using mobile phones (see Section 3.1). Image-based object recognition is needed for this kind of interaction but different variants require different amounts of memory and processing power. In the following we compare the interaction techniques Point & Shoot, Continuous Pointing, and handheld AR. The aim of the study is to investigate the difference between the interaction techniques.

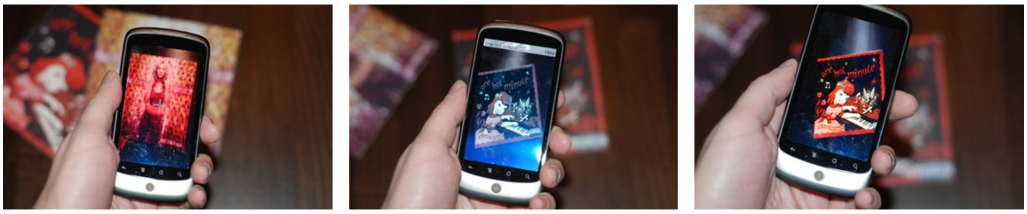


Figure 3.20: Prototypes to compare the three interaction techniques. Point & Shoot (left) shows only the camera image on the screen. The screen must be tapped to trigger the "recognition". Using Continuous Pointing (centre) a CD's title is shown at the screen's top if a CD is recognized. The screen must be tapped to access the detailed view. The handheld AR interface (right) greys out the background. An image of the CD is displayed on top of the object. The user must tap the CD to get further information.

3.5.1 Developed Prototype

To implement the three interaction techniques we aimed at developing a handheld AR system because it has the highest requirements. This enables to derive the realizations of the other interaction techniques from this system. AR systems estimate the pose of the augmenting display in relation to the scene or object that should be augmented. Using this pose a system can transform the augmenting overlay into the reference system of the physical scene and render the augmentation.

In order to mimic a working system we refrained from using visual markers (e.g. QR-codes). In order make handheld AR feasible for a few dozen CDs we extended the approach by Wagner et al. Similar to [HSB09] and the prototype described in Section 3.4 we integrated a Vocabulary Tree in the object recognition pipeline. In the pre-processing phase, images of CD covers are analysed to extract simplified SIFT features [WSB09]. During runtime simplified SIFT features are extracted from the images delivered by the phone's camera and compared to the SIFT features from the CD cover. If the number of matches is above a certain threshold an according homography is computed and used to draw an overlay on top of the camera image. To increase the speed, recognized CDs are tracked (see [WSB09]) in subsequent camera images. We implemented the algorithm for the Android platform using C and Java. The prototype recognizes objects in a 320x240 pixel camera frame with 12 FPS and tracks objects in subsequent frames with about 25 FPS on a Google Nexus One. [WSB09] provides an extensive description of their approach and its performance, considering registration errors, and frame rate. We do not use the same implementation but the performance is similar.

Based on the implemented handheld AR algorithm we designed an application that highlights recognized CDs by greying out the camera image and displaying a coloured image of the CD's cover at the position of the CD. When a user touches the CD a separate view is displayed containing information about the CD and playback controls. Using this system we derived the other interaction techniques. For Point & Shoot the same recognition and tracking algorithms run continuously but hidden from the user. As soon as the user touches the screen to "take a photo" the sound of a single-lens reflex camera is played. If the system has recognized a CD the same but static view as for handheld AR is displayed. For the Continuous Pointing approach, a CD's title is show on the screen whenever a CD is recognized and tracked. Figure 3.20 shows the implementation of the three interaction techniques. By using the same algorithms for all three interaction techniques we can rule out the effect of the algorithm. In practise Point & Shoot would have a higher latency, especially using a server-based approach, but might recognize small object more robustly if photos with a high resolution are used.

3.5.2 Evaluation Method

In order to compare the interaction techniques described above, we conducted a user study that is described in the following. We evaluated the three interaction techniques using the described system. The aim of the study was to assess the participants' preferences, their perceived workload using the different techniques, and to collect qualitative feedback. In the experiment participants performed a single task to compare the techniques. A within-subject design with one independent variable resulting in three conditions was used.

We predicted that handheld AR is more usable than the other interaction techniques because it is always clear to the user which CD is currently available. We further assumed that Continuous Pointing and handheld AR are both more usable than Point & Shoot because they provide continuous feedback to the user. Therefore, we assumed that participants' ratings will reflect these differences.

3.5.2.1 Design

The interaction technique is the experiment's independent variable. Using repeated measures we counterbalanced the order of the three conditions to reduce sequence effects. Dependent variables are the task completion time and participants' subjective ratings. The order of the conditions was counterbalanced using a Latin square design, i.e. the order of conditions was different for each of the subjects. The questions were always asked in the same order. We measured the time participants needed to answer the five questions. More importantly, we asked them to fill the "overall reactions to the software" part of the Questionnaire for User Interaction Satisfaction (QUIS) [CDN88] to estimate the perceived satisfaction and the Raw NASA TLX [HS88] (i.e. the NASA TLX without the weighting process) to assess their subjective task load.

3.5.2.2 Participants and Apparatus

We conducted the study with 14 participants, 6 female and 8 male, aged 22-49 (M=32.21, SD=8.33). All subjects have a higher education five used Point & Shoot before with a commercial application but none of them was familiar with the other interaction techniques. The prototype running on a Google Nexus One was used to execute the task. The investigator selected the interaction technique between the tasks. We prepared three sets of CD covers printed on cardboard. We prepared an additional set of CD covers for the introduction.

3.5.2.3 Procedure

After welcoming a participant we explained the purpose and the procedure of the study. Furthermore, we asked for their age and noted down the participant's gender. Prior to starting the task we demonstrate how to use the three interaction techniques. The participants' task was to answer five questions related to the provided CD covers that were printed on cardboard (e.g. "What is the price of the Amy Winehouse album"). To answer a question they had to read the description provided by the system after selecting a CD. After answering a question participants were asked the next question. After completing all questions with one condition they repeated the task with the next condition, a new set of questions, and different CDs. We asked participants to answer the questions as fast as possible.

3.5.3 Results

After conducting the experiment we collected and analysed the data. We found significant differences between all three interaction techniques for the subjective feedback. We did not find significant effects on the time participants needed to complete the tasks. As the independent variable has three levels we used the Bonferroni correction to reduce the significance levels (i.e. a significance level of $.01\overline{66}$).

An ANOVA shows that the selection technique had a significant effect on the average QUIS's "overall reactions to the software" part ($p < .001$). Using handheld AR leads to a higher score ($M = 5.31, SD = 0.82$) compared to Point & Shoot ($M = 4.04, SD = 1.13, p < .01$) or Continuous Pointing ($M = 3.10, SD = 1.01, p < .001$). The score for Point & Shoot is also significantly higher than the score for the Continuous Pointing ($p < .01$). The results for the individual scores are shown in Figure 3.21.

An ANOVA shows that the interaction technique also had a significant effect on the Raw NASA TLX score ($p < .05$). An F-test showed that the results for the three conditions have an unequal variance and we therefore use the according t-test. Comparing the individual conditions (see Figure 3.22) shows that using handheld AR ($M = 35.86, SD = 7.40$) results in a lower score than using Continuous Pointing ($M = 51.43, SD = 14.60, p < .01$). The difference between Point & Shoot ($M = 43.57, SD = 16.44$) and handheld AR ($p = .12$) or Point & Shoot and Continuous Pointing ($p = .19$) are not significant especially considering the corrected significance level.

Average task completion time for the three conditions are, Point & Shoot: $M = 49.96s, SD = 20.61$, Continuous Pointing: $M = 54.77s, SD = 25.82$, and handheld AR $M = 49.80s, SD = 32.09$ but the differences are not significant (ANOVA: $p = 0.85$).

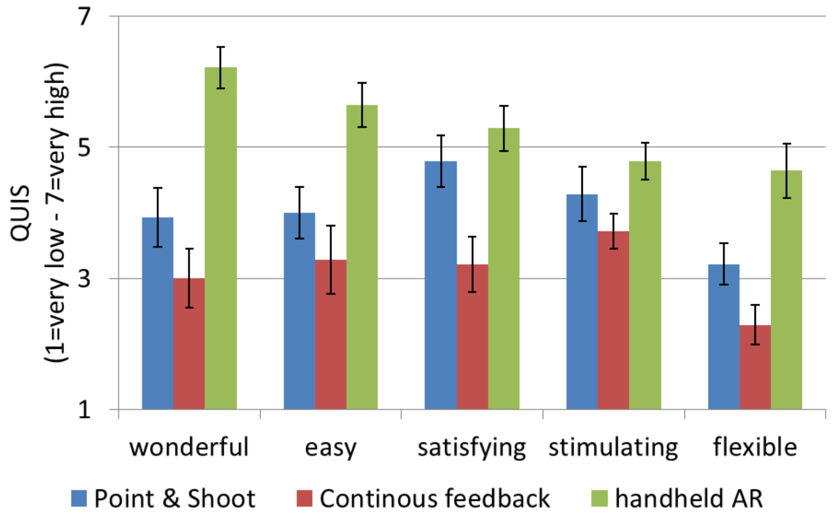


Figure 3.21: Individual scores of QUIS "overall reactions to the software" part for the three camera-based interaction techniques. High values mean positive reactions and low values mean negative reactions (error bars show standard error).

Most of the participants' comments addressed the performance and the accuracy of the object recognition in particular for Continuous Pointing. E.g. six participants criticized the recognition after using Continuous Pointing and said that "the recognition is too slow" or that it "should also work if I hold it [the CD] in my hand". About handheld AR three participants criticized the inaccurate augmentation, e.g. by saying "it's a bit shaky". Participants regularly mentioned that the handheld AR algorithm works better than Continuous Pointing. Participants liked the Point & Shoot recognition but disliked its speed. E.g. one statement about Point & Shoot was that "recognition is best but it's slower" even though there was virtually no latency after triggering the recognition. Participants saw the largest potential in handheld AR. Six participants requested that information should be presented using the augmentation. One participant stated, for example, that information should be visible "instantly on the CD". Participants also stated about handheld AR that "this is cool" or "I want that on my phone".

3.5.4 Discussion

Participants' quantitative and qualitative feedback showed that handheld AR is preferred compared to Point & Shoot but especially compared to Continuous Pointing. The task load shows only small difference between the conditions but handheld AR results in a significantly lower task load than Continuous Pointing. In general the Continuous Pointing interaction technique does not justify the higher requirements compared to Point & Shoot. Participants, however, see much more potential in handheld AR.

The results confirm our expectation that handheld AR outperforms the other interaction techniques and receive better ratings from the participants. The results, however,

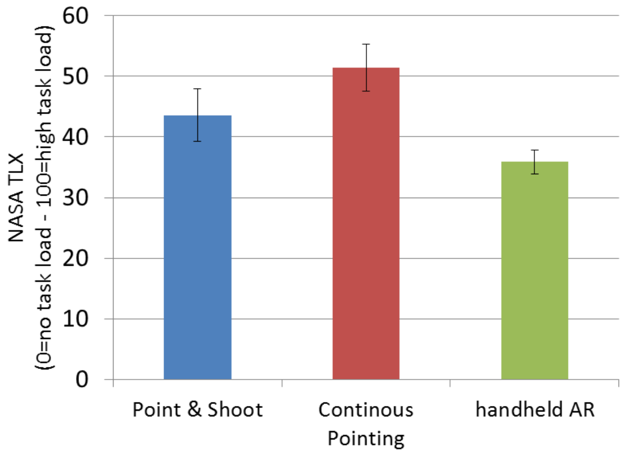


Figure 3.22: NASA TLX scores from 0 (low task load) to 100 (high task load) for the three camera-based interaction techniques (error bars show standard error).

disprove our assumption that Continuous Pointing would outperform Point & Shoot. We assume that the provided feedback is counterproductive. An explicit trigger results in a higher conformity with user expectations. Participants focused on one object at a time before triggering Point & Shoot, similar to taking a photo with a camera. Unconsciously the participants created perfect conditions for the algorithm. While handheld AR permanently communicates the system's status, Continuous Pointing hides the limitation. This might lead to incomprehension and rejection of Continuous Pointing.

The study is limited in a number of ways. The same algorithm is used for all conditions even though if implementing a commercial product one would use algorithms optimized for the particular interaction technique. Furthermore, more male than female participants took part in the study. We collected mostly subjective results and did not found a significant effect on the task completion time. We, however, assume that users' impression and subjective feedback is more relevant for the intended use case but a novelty bias might have affected the participants' ratings of all interaction techniques.

3.6 Summary and Implications

In this chapter we analysed camera-based mobile interaction techniques for accessing information connected to physical objects. In the following we summarize the conducted studies and their results. In addition, we highlight the implications of our findings for the development of mobile camera-based applications for interacting with physical objects.

3.6.1 Summary

In this chapter we analysed and compared four mobile interaction techniques to access information connected to physical objects. We used manual text entry with a soft keyboard as a baseline condition. We compared Point & Shoot with manual text entry in a controlled experiment. Participants were asked to use both interaction techniques to answer questions using advertisement posters as anchors to digital content. Our evaluation with 46 participants shows that the designed interaction outperforms the soft keyboard-based interaction in terms of task completion time and subjective satisfaction. On average, participants are 22.9% faster and assign a 72.3

As we showed that Point & Shoot is a promising interaction technique we conducted an explorative studies to further investigate its characteristics. Ten participants accessed digital images by taking photos of a printed photo book using Point & Shoot. With a recognition rate of 98.3% for all photos that have not been taken to test the system and not wrong results we showed that current algorithms are sufficient for using Point & Shoot under realistic conditions. Conducting the study with a fairly diverse sample, our results indicate that Point & Shoot is not only usable by young early adopters but also by elderly users without a technical background. In order to prepare a comparison of the camera-based interaction techniques we conducted another explorative study to investigate Continuous Pointing. Six participants were asked to use Continuous Pointing for interacting with printed posters. Our results indicate that most users systematically scan posters with multiple annotated regions in zigzag and align the phone with the respective object. Furthermore, by showing that none of the participants was able to find all interactive regions we determined that finding interactive regions using Continuous Pointing without additional can be challenging.

In an controlled experiment we compared the three camera-based interaction techniques. Participants were asked to access information about music CDs. We showed that the participants' overall reaction about handheld AR is significantly more positive than about the other interaction techniques. Participants assigned a 41.6% higher QUIS overall reactions to the software score using handheld AR compared to Continuous Pointing and a 23.9% higher score compared to Point & Shoot. Using handheld AR also lead to a 21.5% lower perceived task load score compared to Continuous Pointing and a 43.4% lower score compared to Point & Shoot. While not significant, other measures also support our conclusion that handheld AR outperforms other camera-based interaction techniques when accessing information about physical objects.

3.6.2 Implications

Based on the series of controlled experiments and explorative studies we can conclude a number of implications for the design and implementation of prospective applications to access information about physical objects.

In total 76 diverse participants that aged between 20 and 70 years took part in the four studies. Among the participants were elderly without experience with mobile phones, students from different disciplines, and experienced smartphone users with a background in computer science. During none of the studies even a single participant was unable to complete the specified task. While this result does not show how usable the tested camera-based interaction techniques are – it very clearly shows that they are at least usable by virtually every user. Furthermore, the very short introduction that participants got before executing the tasks especially in the study described in Section 3.2 indicates that Point & Shoot is intuitive to use.

We showed that the camera-based interaction technique Point & Shoot is better suited to access information about physical objects than a virtual keyboard. This is consistent with previous work by Davies et al. [DCDH05a] even though they could not provide significant evidence. Comparing the three camera-based interaction techniques we also showed that handheld AR is the superior technique. Thus, we can conclude that handheld AR is not only the preferred camera-based technique but also preferred compared to manual text entry using a virtual keyboard. Therefore, we can conclude that handheld AR is a viable option that is at least on a par with established interaction techniques.

Looking at the implementation of different camera-based interaction techniques we found that existing algorithms for Point & Shoot are already adequate. Using a server-based implementation the algorithms are sufficiently robust to be used in various conditions and by diverse users. Successful commercial applications strongly support this conclusion. For Continuous Pointing and handheld AR, however, participants criticize the responsiveness and accuracy of our implementation. Still, at least handheld AR receives better ratings and participants saw the largest potential in this interaction technique. This shows that further work to improve the accuracy and responsiveness of the algorithms and implementations for handheld AR is necessary. Furthermore, the scalability of handheld AR is restricted by the available memory on the phone. Future work must find ways to overcome this dependency.

For the design of the three camera-based interaction techniques we reduced the interfaces to their very core concept. The interfaces of all three interaction techniques presented almost no information unless the participant selected an object. In particular, handheld AR has the potential to provide information directly by the augmentation. Handheld AR could therefore further benefit from a richer design of the augmentation. As no guidance for developing handheld AR interfaces exist today, our results implicate that further research about designing these interfaces is required. Research about the interface design of handheld AR application must address the augmentation of objects inside the camera image as well as the interaction with them. To disburden users from systematically scanning the environment to find objects (as observed in our studies) it is also required to investigate approaches to highlight objects that are not currently inside the camera image. Thus it is necessary to address the design of on-screen content and controls as well as the visualization of off-screen objects.

4 On-Screen Content and Controls in Handheld Augmented Reality

The design of the user interface is crucial when developing interactive systems. This, of course, also applies to handheld AR applications. As we showed in the previous chapter, handheld AR is the most promising camera-based interaction technique to access information connected to physical objects. Handheld AR is, however, still a young field and developing such applications or prototypes requires the use of architectures and algorithms that are currently subject of intense discussion and research. Therefore, handheld AR research is currently dominated by technical development. New algorithms make handheld AR feasible for more and more different types of objects, incremental improvement increases the accuracy, and combining different techniques greatly improved the performance [WS09]. As those technologies are needed to build handheld AR prototypes they are also needed to investigate the interface design of handheld AR systems. Therefore, only the basic characteristics of handheld AR interaction have been studied mainly using abstract tasks (see Chapter 2). The interface design, however, has been mostly neglected so far.

In this chapter we investigate the interface design of handheld AR systems. The goal is to derive general design principles that should be considered when developing such applications. Printed photo books and physical CDs are used as exemplary types of physical media. Design solutions to augment both media types are explored using a participatory approach. Design solutions proposed by participants of two user studies are consolidated and implemented as software prototypes. In two experiments the resulting interface designs are compared to determine their usability. We show that an object-aligned augmentation is more efficient for information presentation. In contrast, it is also shown that input controls should not be presented via an object-aligned presentation. We close the chapter with a summary and an outline of the implications of our findings on the design of prospective handheld AR applications.

4.1 Participatory Design Studies

As there is little work on designing handheld AR user interfaces we decided for an explorative approach to investigate the design space. Following Beaudouin-Lafon and Mackay [BLM02] participatory design is used to generate ideas. We focus on the step Svanaes and Seland call "Tool Making" [SS04] as we are interested in prototypes that are not restricted by current technical limitation. Svanaes and Seland suggest that "participants design specific solutions to specific needs and do not need to worry about issues such as software architecture, implementation, information structure, interface consistency, and integration with other ICT systems". In order to ensure "that everyone contributes, not just those who are verbally dominant" [BLM02] participants are asked individually to develop low-fi paper prototypes of a handheld AR system.

In order to design the user interfaces participants need concrete use cases that they understand. Therefore, we selected two different use cases and conducted participatory design sessions for each of them. The first selected use case is the augmentation of printed photo books with additional information that cannot be reasonably printed on paper. This media type has been selected as a representative for objects that are composed of multiple entities – in this case different photos presented on a single page of the photo book. The second use case is the augmentation of music CDs that should serve as anchors to further information and services. Music CDs are an example of media that can stand alone. The studies for the two types of media have been conducted separately to avoid an influence of one use case on the other one. In the following we first describe the participatory study for printed photo books and afterwards the study for augmenting music CDs.

4.1.1 Interaction with Printed Photo Books

The aim of this study was to collect features and proposals for the interface design from participants. Therefore, we asked participants to specify information and features they want to access using their printed photo book. More importantly, we asked them to propose designs to visualize information with handheld AR using pen and paper.

4.1.1.1 Method of the Study

Participatory design was used as the method for the study. The participants' first task was to specify information and features they consider important for a handheld AR application that augments printed photo books. The second task was to draw design sketches using pen and paper on provided sheets of paper each with an image of a phone that shows a photo book on its display.

12 persons (8 male) participated in the study. The participants' age was between 8 and 54 years (M=32.58, SD=12.35). 4 participants had a technical background (undergraduate and graduate students) and 8 had no technical background.

The provided sheets of paper contained a printed image of a mobile phone that shows an unaugmented image of a photo book on its screen. One of the provided printouts is shown in Figure 4.1. The layout of the used photo books is consistent with the sparse knowledge gained from analysing photo books [SB09] and is also consistent with image composition algorithms for photo books [Atk08].

In the beginning of the study we introduced the participants to the study's purpose and collected demographic data. The remainder of the study was split into the two tasks. In the first part we asked participants to write down a list of information and features they consider relevant. Participants were also asked to rate each of the named features on a five point Likert scale (from not important to very important). Afterwards, we asked the participants to draw an augmentation on the provided image of a mobile phone.

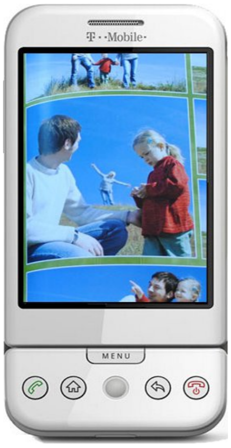


Figure 4.1: Image of a mobile phone that shows an unaugmented photo book on the screen provided to the participants. The image was printed on A4 with the phone scaled to its true size.

Table 4.1: Frequency and rating of the information for an augmentation of printed photo books named by participants.

Information	Mentioned	Rating
persons' names	9	4.2
recording time	8	4.3
recording date	7	4.7
recording place	6	3.8
title/description	6	3.3
object description	5	3.4
comments	4	3.5
tags/categories	3	4.0
related images	3	3.5
links/social networks	3	3.0

4.1.1.2 Results

In the following we report the results from the user study. First the results for the desired functions and information are described, followed by an outline of the sketched interfaces.

Information and functions

The participants named 6.83 (SD=3.43) different information/features on average. We normalized the results by merging synonyms and very similar answers. Table 4.1 shows all aspects that have been mentioned more than two times and their average rating.

The results can be further reduced considering that recording time and date is very similar information that is often presented side-by-side. Persons' names and object descriptions are also similar information that describes specific parts of a photo. Eight participants proposed to not only display information but requested the possibility to also create or change additional content. In particular, participants wanted to add descriptions of persons, sights and other objects to photos similar to the way photos can be annotated in online galleries on Flickr.com and Facebook.

Visualization

We collected 24 sketches (M=2.00, SD=0.60) from the participants. Analyzing the sketches we found that the visualizations can be differentiated by the way the augmentation is aligned. Six participants aligned the information to the border of the phone. Figure 4.2 (left) shows an example of a sketch where the information is presented at the phone's border. The information is located at either one or two sides of the display. All but one of the sketches used the top and the bottom of the screen. In contrast, three participants aligned all information to the photo. Figure 4.2 (right) shows one of these sketches that presents the information at the photo's border. Some participants choose a mixed design where some information is aligned to the phone's border and other information is aligned to the photo itself (see Figure 4.3). Particularly, information that describes only parts of a photo is aligned to this part while general information about a photo, such as its title, is located at the border of the phone's display.

Ten participants explicitly suggested highlighting the recognized photos of the photo book in some way. Seven participants proposed to draw a rectangle around the photos. Other participants suggested to grey out the background or did not specify a particular way. Even though not requested, six participants proposed to have a way to activate additional functionalities on a separate view. Two participants proposed using the phone's menu button and two proposed to use icons (e.g. a video icon) that lead to the separated view. The other two participants did not specify a particular way to activate the additional functionalities.

4.1.1.3 Discussion

The information that has been requested most often by the participants describes the photos' content such as, persons and objects that have been photographed. When and where a photo has been taken is almost equally important. More general textual descriptions of a photo such as tags and a title rank third. Participants could also envisage including social features such as comments or pointer to social networks. In general, it can be differentiated between information that describe particular parts of a photo (e.g. the name of a person), information that describes the whole photo (e.g. recording time or a title), and content provided by other users (e.g. comments). Most participants do not only want to view information but also want to add information. They propose, for example, to be able to select regions of a photo to tag persons, sights, and other objects.

The sketches for the visualization of the augmentation produced by the participants revealed two different patterns. Participants align the elements either to the augmented

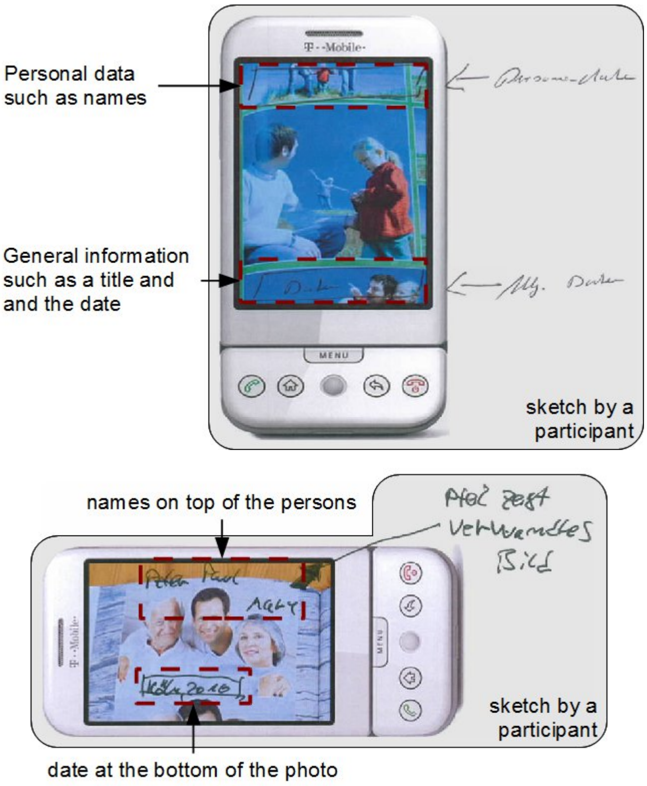

Personal data such as names
General information such as a title and and the date
Personen-daten
Allg. Daten
sketch by a participant
names on top of the persons
Peter Paul
Anke
Köln 2010
Pfeil zeigt verwandtes Bild
sketch by a participant
date at the bottom of the photo

Figure 4.2: Sketches of handheld AR interfaces to augment printed photo books created by one of the participants. Information is either aligned to the top and bottom of the phone's display (top) or aligned to the photo (bottom). For illustrative purpose we selected a very simple sketch.

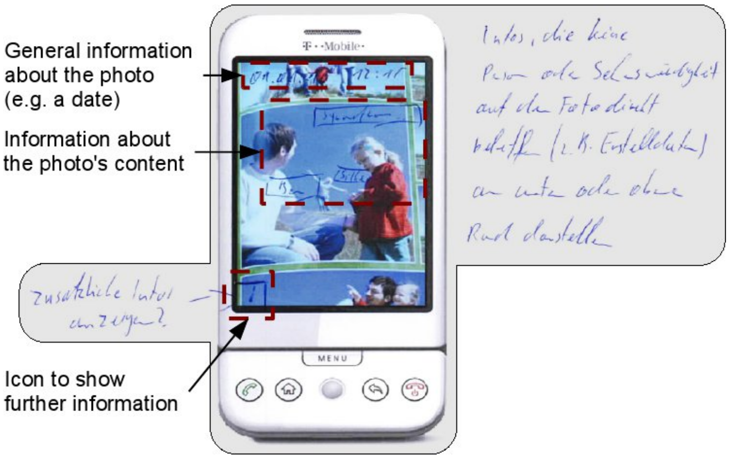

General information about the photo (e.g. a date)
Information about the photo's content
Infos, die keine Person oder Sehenswürdigkeit auf dem Foto direkt betreffen (z.B. Erstelldatum) am unten oder oben Rand darstellen
Zusätzliche Infos anzeigen?
Icon to show further information

Figure 4.3: A sketch of a handheld AR interfaces that has some information aligned to the top and bottom of the phone's display and other information aligned to the photo.

object or they align them to the phone's border. Furthermore, some participants propose mixed designs. Most participants propose to highlight the augmented object.

4.1.2 Interaction with CD

The aim of the second participatory study was the same as for the first study but for a different type of physical media to validate the collected results. We wanted to collect features and proposals for the interface design of a handheld AR application that augments music CDs with additional information from the participants. Therefore, we first asked them to specify the information and features they want to access using their physical CDs. Furthermore, we asked for designs proposals that visualize information with handheld AR using pen and paper.

4.1.2.1 Method of the study

Again participatory design was used as the method for the study. The participants' first task was to specify information and features they consider important for a handheld AR application that augments physical CDs. The second task was to draw a design sketch using pen and paper on a provided sheet of paper with an image of a phone that shows a music CD on its display.

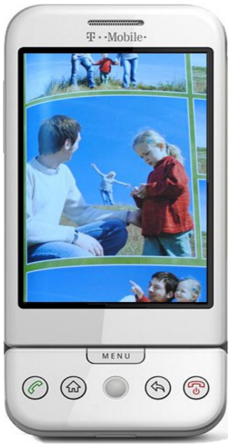


Figure 4.4: Image of a mobile phone that shows unaugmented music CDs on the screen provided to the participants. The size of the sheet is A4 and the printed devices are slightly larger than the real device.

We conducted the study with the help of 11 participants, 2 female and 9 male, aged 22-30 years (M=25.73, SD=2.65). None of the participant took part in the first participatory study. All but one of them are students with a background in computer science. The provided sheets of paper contained a printed image of a mobile phone that shows an unaugmented image of one or two CDs on its screen. One of the provided printouts that contains two CDs is shown in Figure 4.4. The printed device is slightly larger than the device's true size.

In the beginning of the study we introduced the participants to the study's purpose and collected demographic data. The remainder of the study was split into the two tasks. In the first part we asked participants to write down a list of information and features they consider relevant. Afterwards, we asked the participants to sketch a graphical user interface for a handheld AR system that augments CDs. Participants were asked to consider that the device has a touchscreen and we briefly explained the concept of handheld AR. They were free to use multiple sheets of paper to sketch different ideas or discard drafts.

4.1.2.2 Results

In the following we report the results of the study. First we summarize the desired functions and information that should be displayed and afterwards an overview about the layouts of the collected sketches is provided.

Information and functions

On average participants named 4.27 (SD=1.56) functions or information. We normalized the results by merging synonyms and very similar answers. Table 4.2 shows functions and information that has been mentioned more than two times.

Table 4.2: Frequency and rating of the information and functions for an augmentation of music CDs named by participants.

Information	Mentioned
Playback	9
Basic information	7
Wikipedia article	5
Buy music	4
Buy related products	3
Video	3

Nine out of eleven participants requested playback controls to listen to the music that is stored on the respective CD. Seven participants proposed to show basic information about the music CD that includes the CD's title and the band's name. The CD's Wikipedia article that provides more detailed information is requested by five participants. A function to buy the music stored on the CD has been requested by four participants and another three subjects requested a function to buy related products (e.g. concert tickets). Providing related videos has been requested by three participants.

Visualization

We collected 16 sketches from the participants in total. In general, the sketches are diverse and some are rather unorthodox (see Figure 4.5 for an example). After revising all sketches we identified that sketches can be classified in three dimensions. We classify the sketches by the way playback controls, services, and information are presented, how the recognized CDs are highlighted, and what user input is needed to access services.

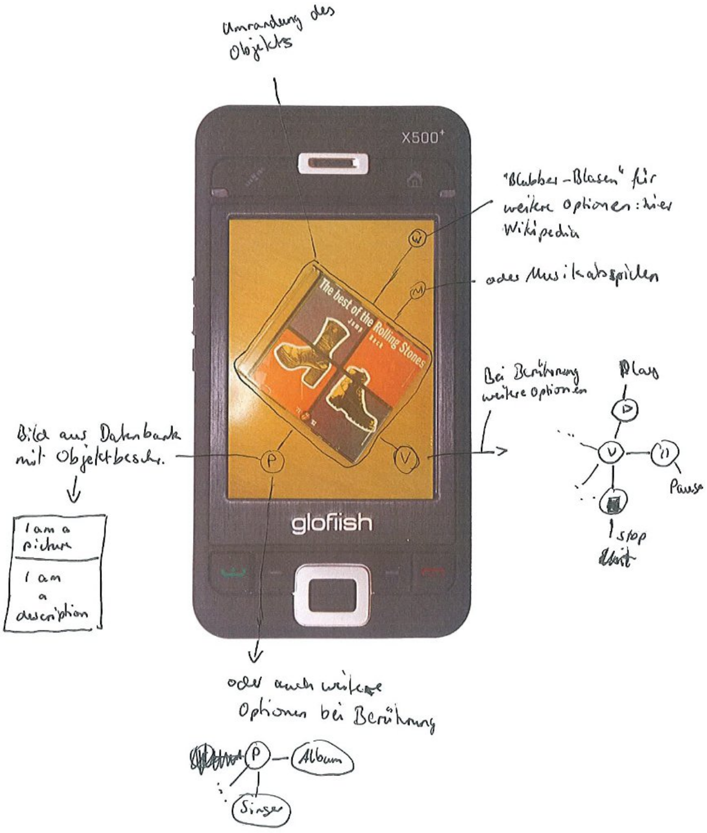

Umrandung des Objekts
X500+
"Blubber-Blasen" für weitere Optionen: hier Wikipedia
W
oder Musikabspielen
M
The best of the Rolling Stones
Bei Berührung weitere Optionen
Play
Pause
stop
Exit
V
Bild aus Datenbank mit Objektbeschr.
P
I am a picture
I am a description
glofiish
oder auch weitere Optionen bei Berührung
P
Album
Singer

Figure 4.5: A rather unorthodox interface to augment music CDs sketched by one of the participants. In this example, bubbles connected with the music CD float around the screen.

Playback controls, services, and information can be presented aligned with the phone (on top of the camera image) or aligned with the CD (inside the camera image). Participants designed solutions using both approaches and, in addition, hybrid solutions that present some information aligned with the phone and others aligned with the CD. Sketches contained icons that represent the availability of some sort of service (e.g. 'W' for Wikipedia). Touching one of these icons invokes the respective functionality. If sketches contained text with the CD's metadata (e.g. title of an album and the year of its release) the text was usually aligned with the phone. Only one participant aligned text with the CD.

Participants designed two approaches to highlight the recognized CDs. Highlighting recognized CDs was proposed for sketches that presented information aligned with the phone, in particular. Three participants suggested that the camera image should be presented in greyscale and only the recognized CDs should be coloured. Three different participants proposed to draw a coloured rectangle around the respective CDs.

For most sketches the functionalities are accessible with a single touch. This was achieved by directly connecting a function with a touchable icon. E.g. a typical play button directly attached to the top of the CD together with other playback controls. A two-stage process was proposed by two participants for selecting a CD if multiple CDs are recognized at a time. This was considered necessary if the icons are not aligned with the CD but with the phone. Two participants intend to make some functionality (e.g. a link to YouTube) accessible using a menu that pops up if the user touches a menu icon.

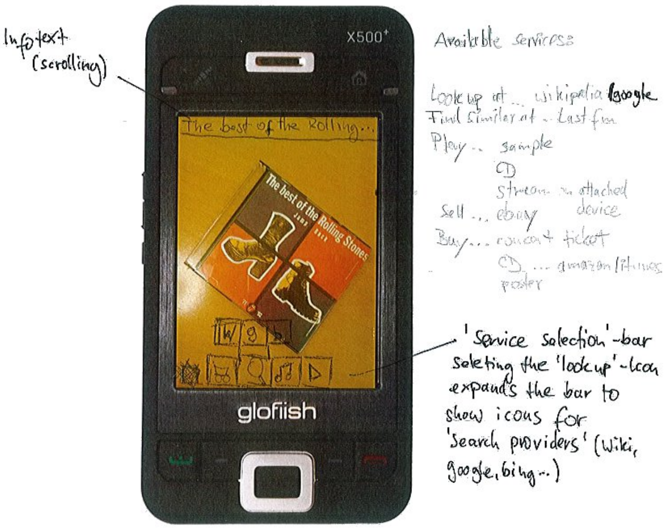


Figure 4.6: Design proposal for augmenting music CDs produced by one of the participants. Information and controls are aligned to the borders of the phone's display.

4.1.2.3 Discussion

The functionality that has been requested most often is to play the music stored on the augmented CD and three participants also asked to provide according videos. Most frequently the participants requested basic information but some also asked for more detailed information in the form of a Wikipedia article. Another area of interest is to either buy related products or sell the CD. In particular the playback function need further investigation because a complete set of functions, such as play, stop and pause, is needed to control music playback (see Section 4.2.2).

Similar to the results of the study for photo books, participants aligned the augmentation to the object or to the phone's display. In addition to the first study, some participants proposed an approach that consists of two steps. Information and functions are only available after a CD has been selected. Again, most participants proposed to highlight the augmented object. While not all participants proposed the highlighting explicitly we assume that this a general demand.

4.2 Interface Designs

The interface designs proposed by participants of both studies can be divided in those were the placement of information and controls is aligned to the phone and those were the information is aligned to the object. Furthermore, some participants proposed a mixture of both approaches. We decided to not design a mixed augmentation in order to investigate the alignment aspect without the ambiguity of a mixed design. For augmenting CDs participants also proposed a two-step interaction that requires selecting the CD before accessing information or functions.

Participants suggested to highlight photos and annotated regions for the printed photo books as well as for the music CDs. They proposed to either draw a rectangle around the augmented object or to grey out the background. Therefore, we decided to combine both approaches. The pages of the photo book and the music CDs are highlighted by displaying only the respective object with colours and leaving the surrounding greyed out. Furthermore, individual photos and annotated regions of a photo are highlighted by drawing a rectangle around them. The centre of the display is marked with a crosshair.

In the following we describe the interface designs derived from the participatory study for augmenting printed photo books as well as for augmenting music CDs. First, we will describe the designs that align interface components to the respective object and afterwards the designs that align the interface components to the phone's display. A further design that requires selecting the object before using it is only developed for music CDs because no participant requested such an approach for augmenting photo books.

4.2.1 Information and Functions

To design concrete interfaces it is necessary to identify which information to present and which functions should be available. The small screen size of mobile phones limits the amount of interface components that can be presented concurrently. Furthermore, handheld AR is mainly useful for browsing information using the physical world as an anchor for further content. The interface should therefore allow quick browsing of information and further details might be provided on separate views. Furthermore, the camera image must be shown on the screen which reduces the screen space available for other interface components. Therefore, we decided to reduce the amount of information and functions that is directly accessible to the minimum. Further information and functions should, however, be provided on a separate view that is accessible by selecting the object.

For augmenting printed photo books nine out of twelve participants requested to show persons' names. Similar information such as a photo's title and descriptions of objects

Table 4.3: Frequency and rating of functions to control music playback.

Function	Importance	SD
Next track	4.7	0.2
Start	4.5	0.5
Increase volume	4.2	0.6
Decrease volume	4.0	0.9
Previous track	4.0	0.7
Pause	3.6	2.3
Stop	3.3	1.3
Repeat	3.2	2.2
Wind back	2.4	2.3
Fast forward	2.1	1.7
Shuffle	2.1	2.5

have also been requested five and six times. Therefore, we decided to present two types of information: A title that conveys general information about a photo and descriptions of regions in the photo. These regions can be persons and other types of objects such as sights. A photo's recording time and place is not directly visible because we assume that this information might not be important for all photos. In particular, if multiple photos, that are presented side-by-side, have been recorded within a close time frame and location. If recording time and location is important for a photo the information can, however, be included in the photo's general description.

To augment music CDs nine out of eleven participants asked for the possibility to control music playback. Furthermore, seven participants requested to have basic information about the CD available. Therefore, we decided to provide functions to control music playback and general information about the CD consisting of the CD's title, the band, and the release year. Providing more detailed information in the form of a Wikipedia article, suggested by five participants, require large parts of the screen. Therefore Wikipedia articles cannot be made directly accessible but could be provided on a separate view.

4.2.2 Function Set to Control Music Playback

Playback controls are considered as the most important feature for an handheld AR system that augments music CDs. In order to provide these playback controls it is necessary to decide which functions are required. Even though mobile music player are almost pervasive we found little work that shows which functions are how important. Thus we asked potential users to determine the most important functions to control music playback. 11 functions were derived from the user interface of common digital music player and mobile music player. We asked 10 participants (5 female, M=30) to rate the importance of each of these 11 functions for the intended use case on a 5 point Likert scale. The results are outlined in Table 4.3.

Most results, e.g. that start is rated more important than stop, are not surprising. The outcome is consistent with a function-set that Kranz et al. [KFH$^+$06] derived from a user study (even though, it is not completely clear how they gained their results) and the function-set we determined in our own work for a gestural music player [HLB$^+$10].

To derive the final function set it is necessary to consider that most functions come in pairs. E.g. a function that increases the volume is only reasonable if a function that decreases the volume is also available and start must be accompanied by a stop. We decided to support the seven highest rated functions which makes repeat the highest rated function that is not included.

4.2.3 Object-Aligned Interface Components

The object-aligned interfaces connect the interface components to the augmented objects. I.e. the interface component follows the movement of the object inside the camera's video.

4.2.3.1 Photo Books

The interface design to augment printed photo books, shown in Figure 4.7, attaches the information to the photos. A black border frames the photo to highlight it. The title is shown aligned with the top of the photo's frame. It is displayed above the photo to not occlude the photo. The text is displayed in black with a white border around the characters to ensure a high contrast and improve readability.

Regions of the photo are also highlighted by drawing rectangles around them. A white border is used to make the frame that highlights a photo and regions easily distinguishable. A region's description (e.g. a person's name) is shown above or below the region aligned with the rectangle's position and orientation. The augmentation follows the movement of the photo inside the camera's video. If multiple photos are visible all photos are highlighted and each has its own elements visible simultaneously.

4.2.3.2 Music CDs

For the object-aligned interface to augment music CDs the buttons to control music playback and the general information about the CD must be presented. The interface, shown in Figure 4.8, attaches both to the CD. The playback controls are aligned with the bottom of the CD to make them easier to touch if the CD is oriented upright inside the camera image. General information is aligned with the top of the object. In order to ensure a high contrast the text is shown with black characters. The background of the text is coloured with a semi-transparent white to increase the contrast.

Six buttons are used for the functions to control music playback. Icons from standard music players are used to show a button's function. As hiding parts of the object is less critical than for photos buttons and text is shown on the CD and not below or above. The

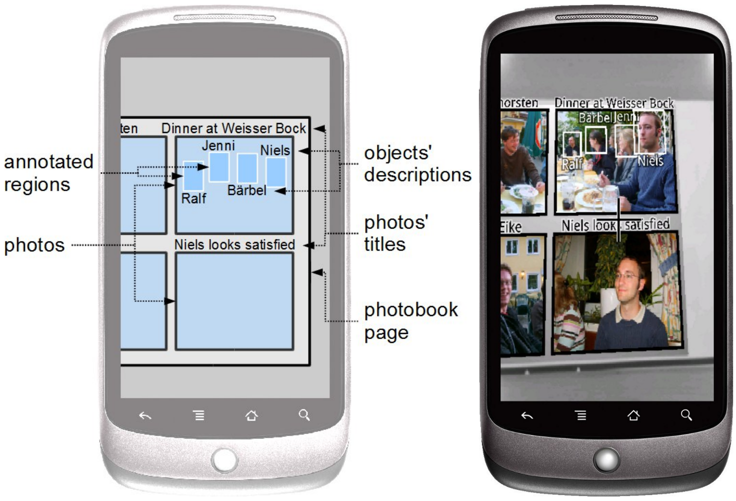


Figure 4.7: Concept and mock-up of the object-aligned interface to augment photo books. The design aligns the annotations to the photos and thus the annotation is visible for all photos simultaneously.

overlay follows the movement of the CD inside the camera's video. If multiple recognized CDs are visible all CDs are highlighted and each has its own control elements.

4.2.3.3 Discussion

The general advantage of the object-aligned interfaces is that multiple objects are accessible at the same time. The information about different photos on one page, for example, is visible concurrently. Furthermore the association between interface elements and objects is unambiguous. A CD's input controls are directly attached to the CD and it is obvious to which CD a button belongs or which title belongs to which photo. The size of the presented text can naturally be increased and decreased by changing the distance of the phone to the object. Therefore, screen space is used efficiently. In addition, this design is rotation invariant. As the orientation of text is aligned to the augmented object the user can hold the phone upright or sideways.

A disadvantage of these interface designs is that the augmentation becomes small if the phone is far from the photo. Thereby, text becomes difficult to read if a user wants to get an overview about the information available for a photo book page or multiple CDs. Furthermore, the text permanently moves and wobbles if the user moves the phone. Thereby, readability of text is affected by accidental movement of the phone. Adding seven physical buttons to a CD would result in a button size of almost 2cm. Thus, the size of augmenting buttons is fairly large in the physical domain. Depending on the distance between the phone and the CD the size of the virtual button is, however, much smaller if the CD does not fill the entire screen. Therefore, this interface for CDs has the additional disadvantage that touching the buttons is also affected by accidental movement of the phone.

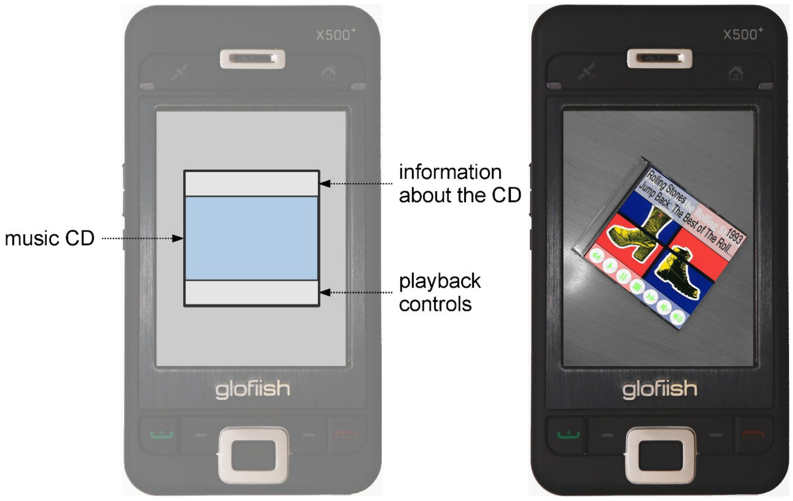


Figure 4.8: Concept and mock-up of the object-aligned interface to augment music CDs. The design aligns the buttons to control music playback to the CD's bottom and information about the CD to the CD's top.

4.2.4 Phone-Aligned Interface Components

The phone-aligned interface aligns the interface components with the phone's display. That means that text and buttons are always at the same position of the display and do not follow the movement of the object inside the camera's video.

4.2.4.1 Photo Books

The second interface design for photo books, shown in Figure 4.9, aligns the information to the phone. The photo's title is shown at the top of the screen. The text is displayed in black with a white semi-transparent background. As only one title can be displayed at a time the photos must be selected first. In order to make this selection fast and easy to use the photo below a crosshair in the centre of the screen is automatically selected. Information is only displayed for the photo that is located below the crosshair. To select a photo the user has to move the crosshair over the photo by changing the phone's orientation or position. Thus, even if multiple photos are visible in the camera image only the information for one photo is displayed. The same black border as for the object-aligned interface frames the selected photo to highlight it.

Regions of a selected photo are highlighted by the same white rectangle used for the object-aligned interface. The textual description of a region is shown at the bottom of the screen. The text is also displayed in black with a white semi-transparent background. If a photo contains multiple annotated regions the according text is placed in a continuous row. To connect the regions with their textual description a white line is drawn from the border of the region to the white box that frames the textual description.

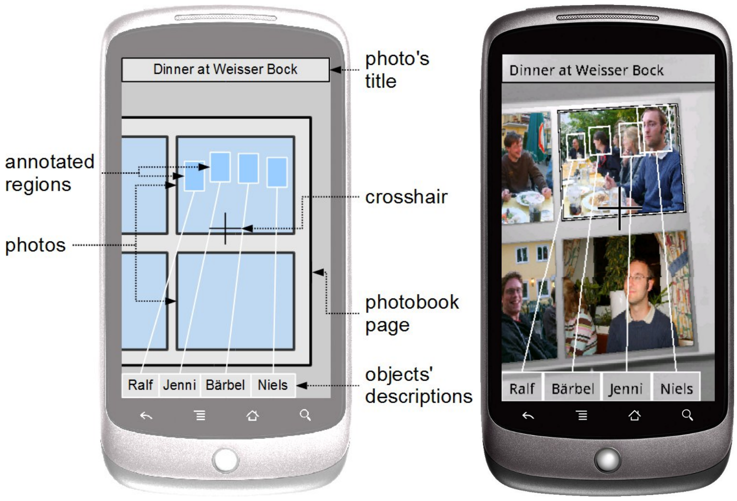


Figure 4.9: Concept and mock-up of the object-aligned interface to augment photo books. The design aligns the annotations to display's top and bottom and only the annotations of the photo below the crosshair are visible.

4.2.4.2 Music CDs

The phone-aligned interface design for music CDs shown in Figure 4.10 aligns the controls to the phone. Similar to the interface for photo books the general information is aligned to the top of the screen. The text is displayed in black with a white semi-transparent background.

The playback controls are displayed at the bottom of the screen. The same buttons and icons as for the object-aligned interface are used. The icons are always at the same position and fade to grey when no CD is below the crosshair. If multiple CDs are found in the camera's video only the most prominent CD (the one that takes most screen space) is highlighted.

4.2.4.3 Discussion

The advantage of this design is that interface components remain at a fixed position. Thereby, text has always the same readable size and stays at the same position as long as an object remains selected. Readability is not affected by changing the phone's position. Since the buttons for music CDs do not change position movement of the phone does not affect the interaction with the buttons. Furthermore, this approach has the advantage that the buttons are relatively large.

The designs' disadvantage is that the interface components for only one object are available at a time. For the augmentation of photo books this means that the user must move the crosshair across all photos on a photo book page to get an overview about the available information. Similarly only one CD is accessible at a time and it might not always be apparent which CD is currently augmented.

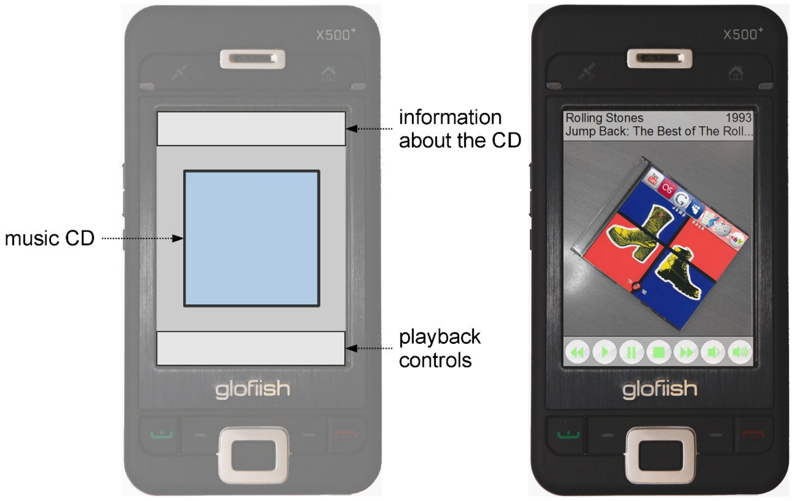


Figure 4.10: Concept and mock-up of the phone-aligned interface to augment music CDs. The design aligns the buttons to control music playback to the display's bottom and information about the CD to the display's top.

4.2.5 Two-Step Interface for Music CDs

Participants of the user study proposed a two-step interface as an additional approach for the interaction with music CDs. The derived design (see Figure 4.11) does not make information and interaction controls directly available. A CD must first be explicitly selected by touching it. Information and functions are only accessible on a separate view that is shown after a CD has been selected. To make this interface comparable with others the same general information and playback controls as for the other designs are used on the separate view. If multiple CDs are visible at the same time all are highlighted in the same way and each of them can be selected by touching the CD.

The design has the advantage that it consumes the least screen space and does not alter the camera's video beyond greying out the background. The clear drawback is that a CD must be selected before accessing any functionality, however, while interacting with a CD the CD does not need to be in the focus of the camera. Thus, the user does not need to hold the phone over the CD while using the playback controls. Furthermore, additional information could be made available on the separate view compared to the other designs.

4.2.6 Selection Techniques for Regions of Photos

As annotating regions of photos has been requested in the participatory study we designed three selection techniques to mark regions. As we did not collect recommendation for designing this interaction from participants we designed three fundamentally different approaches inspired by previous work. With the first two selection techniques

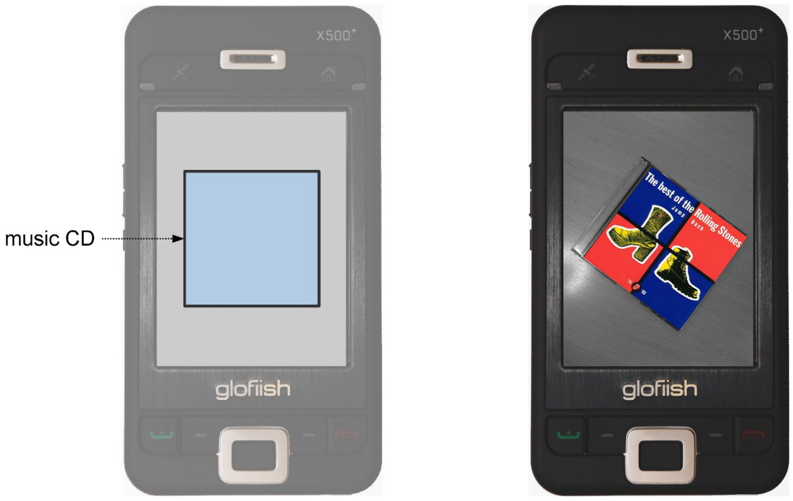


Figure 4.11: Concept and mock-up of the interface design for augmenting music CD that requires selecting the CD before using it.

users select regions in the reference system of the augmentation. They either have to move the phone or touch on the display. We included a third technique where regions are marked by touching a separate static image as a baseline. The three techniques are shown in Figure 4.14. We did not include the selection techniques Liao et al. proposed with the PACER system [LLLW10] because we aim at true handheld AR instead of loose registration and we cannot exploit knowledge about distinct document regions (e.g. words and sentences). The two touch-based techniques can, however, be seen as the basic concepts that are combined in PACER.

4.2.6.1 Crosshair-Based Region Selection

With the first technique, outlined in Figure 4.12, the user aims with the crosshair that is located in the centre of the display at a corner of the region that should be selected. The technique is inspired by handheld AR systems that use a crosshair in the centre of the screen to select predefined objects (e.g. [RO08]). By touching the display at any position the user defines the first corner (e.g. the top-left corner) of the region. The user than has to move the crosshair to the opposite corner (e.g. the bottom-right corner) by physically moving the phone while touching the display. The region is marked when the user stops touching the screen. As the region is created in the reference system of the augmentation the created rectangle is aligned to the photo.

The advantage of this technique is that the "fat-finger problem" [SRC05] (i.e. that users using touchscreens occlude the area they want to touch) is avoided. As the location at which the crosshair aims can be estimated more precisely than the position in which a finger touch results it might also be more precise. As the user can zoom (by changing the distance of the phone to the photo) while moving the crosshair, the region can be created precisely. Furthermore, the interaction technique can be used with one hand because

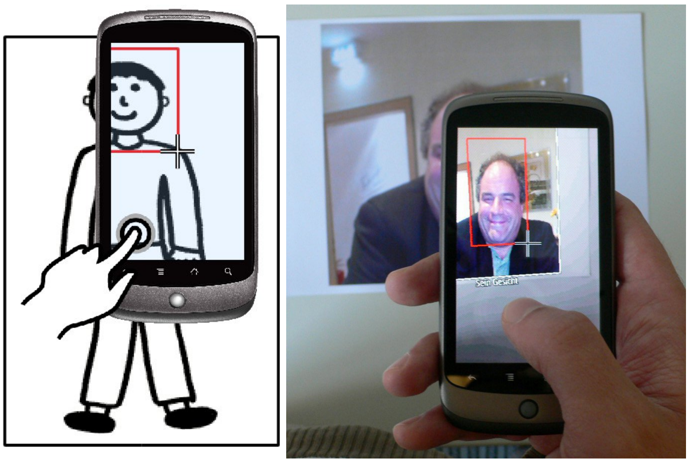


Figure 4.12: The crosshair-based techniques to select regions of photos.

it is not important where to touch the display. While holding the phone in one hand the display can still be easily pressed with this hand's thumb. A disadvantage is that the device must be physically moved. This could lead to a higher physical and mental demand and makes the technique prone to accidently movement of the phone.

4.2.6.2 Augmented Touch-Based Region Selection

With the second technique, shown in Figure 4.13, the user touches at a corner of the region that should be marked to define the first corner. Afterwards, the user has to move the finger to the opposite corner. The region is marked when the user lifts the finger from the screen. As the region is created in the reference system of the augmentation the created rectangle is aligned to the photo even if the user rotates the phone. This technique is inspired by handheld AR systems where users have to touch the augmentation of predefined objects to select them (e.g. [HB10b]).

The advantage of this interaction technique is that the user does not has to physically move the phone. Therefore, it might be less physically and mentally demanding. The user can, however, imitate the crosshair-based approach by not moving the finger but only the phone. In this case the techniques have almost the same advantages and disadvantages as the crosshair-based approach. In any case, however, the technique is affected by the "fat-finger problem".

4.2.6.3 Unaugmented Touch-Based Region Selection

The third technique that is shown in Figure 4.14 serves as a baseline that works on a static image of the photo that should be annotated. As with the previous technique the user touches at a corner of the region that should be marked to define the first corner. The user then has to move the finger to the opposite corner. The region is marked when the finger is lifted from the screen. We did not use more sophisticate selection techniques

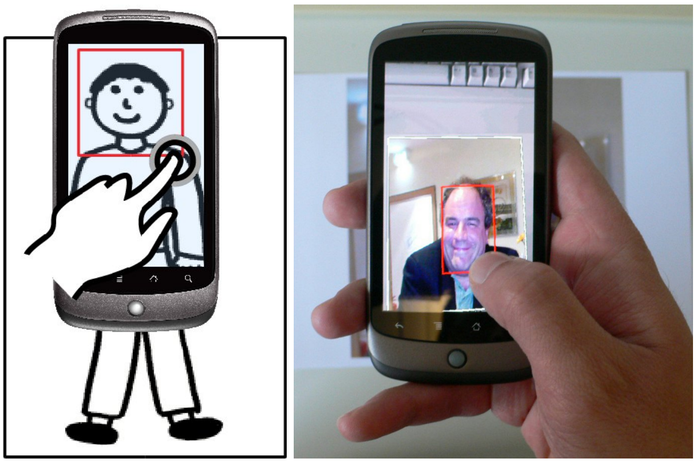


Figure 4.13: The interaction techniques to select regions of photos that requires to touch in the augmentation.

because the other techniques would benefit similarly from more sophisticated interaction techniques.

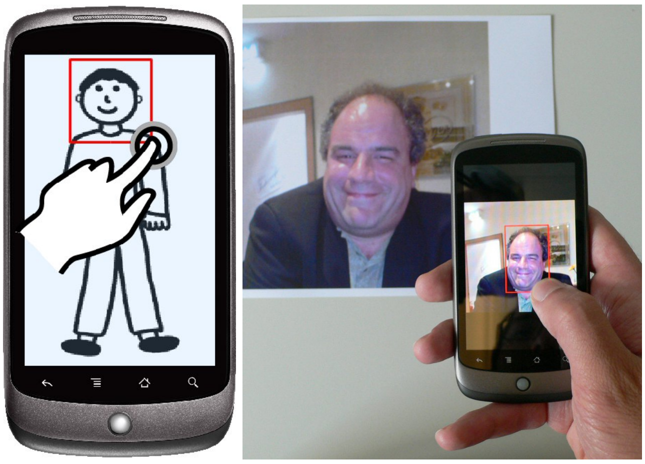


Figure 4.14: The interaction technique to select regions of photos that uses a static picture.

The clear advantage of this technique is that the user can freely move the phone as the phone is disconnected from the physical photo. Thus, unintentional jitter is avoided. Using a separated view can, however, be an imminent disadvantage for a handheld AR system because it requires to switch the view. Another limitation is that the user occludes the area where she aims at with the finger. For this concrete design (but not generally) a further limitation is the lack of zoom. We intentionally did not provide zooming to provide a simple way of interaction that enabled one handed usage.

4.3 Evaluation of the Design Alternatives for Photo Books

Two interface designs to augment photo books have been derived from the participatory study and three techniques to select regions of photos have been developed. In order to investigate these visualization and selection techniques developed in the previous section, we conducted a user study. The aim of the study was to determine differences between the visualizations and the selection techniques.

4.3.1 Method of the Study

In the controlled experiment participants performed one task to compare the visualizations and one task to compare the selection techniques. A within-subject design with one independent variable (two conditions in the first task and three conditions in the second task) was used for both tasks.

4.3.1.1 Design

The study is an experiment with a within-subject design that consists of two tasks. In the first task the independent variable is the interface design consisting of two conditions. The first condition is the object-aligned interface and the second condition is the phone-aligned interface. The selection technique is the independent variable of the second task resulting in three conditions. The order of the conditions was counterbalanced to reduce sequence effects.

4.3.1.2 Participants and Apparatus

We conducted the user study with 14 participants, 6 female and 8 male, aged 23-55 years (M=31.21, SD=8.6). Five subjects had a technical background (mostly undergraduate students) none of them was familiar with handheld AR.

The prototype described in Section 3.5.1 running on a Google Nexus One was used for both tasks. The investigator selected the visualization and interaction technique between the tasks. For the first tasks we prepared two photo books printed on A4 and annotated each of the containing photos with a title and/or regions of the photo describing parts of it. The theme of the first photo book was a wedding and the theme of the second photo book was the visit to a fun fair. We prepared an additional photo book for the introduction with photos taken at a scientific conference. Figure 4.15 shows the photo book used for the introduction. For the second task we printed 20 photos on A4.

4.3.1.3 Procedure

After welcoming a participant we explained the purpose and the procedure of the study. Furthermore, we asked for their age and noted down the participant's gender. Prior to each task we demonstrate how to use all conditions.

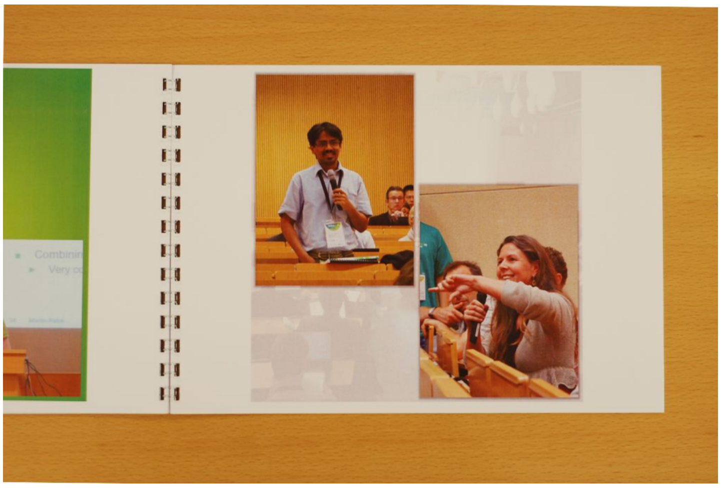


Figure 4.15: Photo book with photos taken at a scientific conference used for the introduction of the user study. The particular page of the photo book contains two photos and an additional photo in the background.

In the first task, participants had to answer five questions related to the photos in the provided photo book. To answer a question they had to read the augmentation shown on the mobile phone. Participants had to combine the information provided by the photos with information provided by the augmentation. E.g. one question was "Who watches soccer?". For this example participants must identify the photo with persons watching soccer and read the annotation that contains the persons' names. After answering a question participants were asked the next question. After completing all questions with one visualization technique they repeated the task with the other visualization and another photo book. We asked participants to answer the questions as fast as possible. The order of the conditions and the order of the used photo book were counterbalanced. We measured the time participants needed to answer the five questions. Furthermore, we asked them to fill the NASA TLX [HS88] to assess their subjective task load and the "overall reactions to the software" part of the Questionnaire for User Interaction Satisfaction (QUIS) [CDN88] to estimate the perceived satisfaction.

In the second task, we asked the participants to select regions on provided photos. With each of the three selection technique the participants had to mark a region in three photos (e.g. "Mark the person's face."). They could repeat marking a region if they were not satisfied with the result. Participants were asked to mark the region as fast and precisely as possible. After completing the task with one selection technique participants repeated the task with the next technique and a new set of photos. The three conditions were counterbalanced to reduce sequence effects. We measured the time needed to mark each region, the coordinates of the region, and how many attempts participants needed. Furthermore, we asked participants to fill the NASA TLX and the "overall reaction" part of the QUIS.

4.3.2 Hypothesis

For the first task we predicted that the photo-aligned presentation is more usable than the phone-aligned presentation. With the photo-aligned presentation the user can see all information simultaneously and can quickly focus on different texts by changing the distance of the phone to the photo book. Therefore, we assumed that participants perceive this condition as less demanding and give it a lower NASA TLX score. Due to the same reasons we assumed that participants would give a higher QUIS score to the photo aligned presentation.

For the second task we assumed that the crosshair-based technique would receive a higher QUIS score and that this condition is perceived as less demanding, which would result in a lower NASA TLX score. We assumed that because, compared to the other conditions, the crosshair-based technique can be used with a single hand and the user can zoom and change the selection simultaneously just by moving the phone. For the touch-based techniques we assumed that unaugmented touch would be more usable because the movement of the hand does not move the image that should be selected.

4.3.3 Results

After conducting the experiment we collected and analyzed the data. We found significant differences between the two visualization techniques as well as between the three selection techniques. We did not find significant effects on the time participants needed to complete the tasks. Participants' qualitative feedback was translated to English.

4.3.3.1 Augmentation Design

Comparing the two visualization techniques we found that the augmentation design had a significant effect ($p < .05, r = 0.81$) on the weighted NASA TLX score (see Figure 4.16). The perceived task load is lower ($M = 103.64$) if the augmentation is aligned to the photo compared to the augmentation that is aligned to the phone's border ($M = 117.86$).

The augmentation design also had a significant effect on the participants average rating of the QUIS's "overall reactions to the software" part ($p < .001, r = 0.70$). On average the rating is higher if the augmentation is aligned to the object ($M = 6.60$) compared to the score for the phone-aligned visualization (M=5.36). The individual scores are shown in Figure 4.17). The visualization technique had a significant effect on the results of all questions ($p < .001$ for the first three questions and $p < .05$ for the others). Task completion time using a photo aligned augmentation is $M = 251s$ ($SD = 95s$) and $M = 270s$ ($SD = 127s$) for the phone aligned augmentation but the difference is not significant (ANOVA: $p = 0.07$).

Most of the participants' comments addressed the performance and the accuracy of the object recognition. E.g. one participants mentioned that "its shaking - probably I

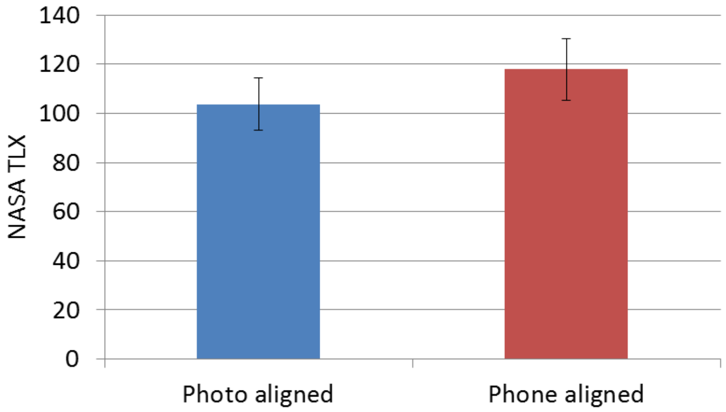


Figure 4.16: NASA TLX for the two interface designs to augment photo books. Low values mean no task load and high values mean high task load (error bars show standard error).

hold the camera wrong" and another participant stated that "the recognition should be faster" and the system "should tolerate bended pages". Participants mentioned for both conditions that the recognition works better than with the other condition.

We observed for both conditions that some participants prefer to hold the phone sideways. This lead to negative comments about the phone-aligned presentation. E.g. "it's difficult to read because the text is skewed" or "have to turn the phone to read the text". About the photo-aligned condition participants mentioned that "it provides a good overview" and "you can see everything". However, they also mentioned that this presentation is "a bit overloaded" and "you have to go near to use the functionality".

4.3.3.2 Selection Techniques

In the second task we compared the three selection techniques. As we used three conditions the significance levels for the follow-up t-tests are reduced to $0.5/3 = .01\overline{66}$ with a Bonferroni correction. The analysis of variance (ANOVA) shows that the selection technique had a significant effect on the weighted NASA TLX score ($p < .01$). Comparing the individual condition (see Figure 4.18) shows that using unaugmented touch ($M = 84.07$) results in a lower score than using augmented touch ($M = 139.36, p < .01$) or the crosshair-based technique ($M = 161.79, p < .001$). The score for augmented touch is lower than for the crosshair-based technique but, considering the corrected significance level, the effect is not significant ($p = 0.025$).

An ANOVA shows that the selection technique also had a significant effect on the average QUIS's "overall reactions to the software" part ($p < .001$). Using unaugmented touch leads to a higher score ($M = 6.57$) compared to augmented touch ($M = 4.97, p < .01$) or the crosshair-based technique ($M = 4.23, p < .001$). The score for the augmented touch technique is also higher than the score of the crosshair-based technique (see the individual scores in Figure 4.19) but without a significant effect

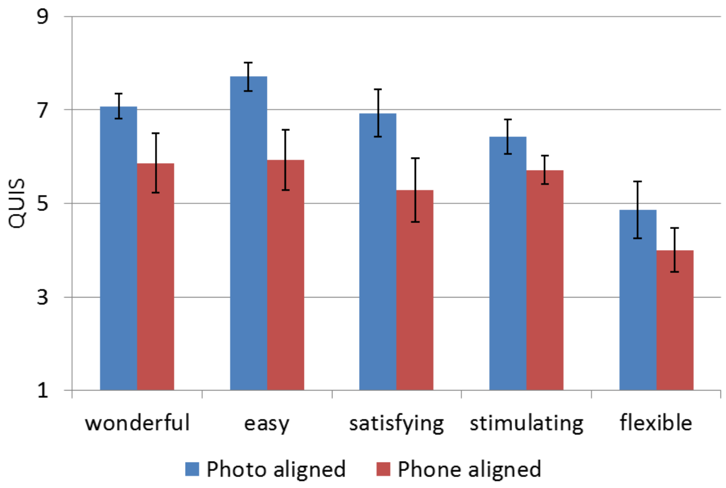


Figure 4.17: QUIS "overall reactions to the software" part for the two interface designs to augment photo books. High values mean positive reactions and low values mean negative reactions (error bars show standard error).

($p = 0.027$). Average task completion time for the selection subtasks are crosshair: $M = 5.4s$ ($SD = 4.3$), augmented touch: $M = 6.8s$ ($SD = 5.1$), and touch $M = 4.1s$ ($SD = 2.2$) but the differences are not significant (ANOVA: $p = 0.08$).

Even though, we demonstrated the techniques prior the task and asked the participant if he/she understands the technique, some participants did not understand the crosshair-based technique. One participants, for example, noted that "it is difficult to touch the crosshair" although it is not necessary to touch it. Mentioned reasons why this condition performs worse than the others are because it is an "unusual interaction" and that it is "difficult to mark a picture by moving the phone". Another participant noted that "moving the whole body is not comfortable". Further comments are that it is "difficult to catch the crosshair where I want it to be" and the same participants stated that "I always forget paying attention to the crosshair". An advantage participants identified is that "the finger does not occlude the object" and that this technique is "usable with one finger".

For the augmented touch technique four participants appreciated that "it has zoom" (compared to the last condition). Compared to the crosshair-based technique they liked that "one can draw the window with the finger". This condition's most often mentioned limitation is that "the device moves when dragging the box" and that "touching changes the position of the phone" or more generally: "it shakes too much for me".

We got mostly positive comments about the unaugmented touch condition. However, participants identified only one advantage of this technique, even though most participants commented on this advantage. They liked that the "image does not move" and that "the image freezed". They also explicitly stated it is "easy to select because it [the image] does not move". The main limitation the participants identified is that "it has no zoom", that "zooming would be nice" and that it is "less precise than the cursor without zoom".

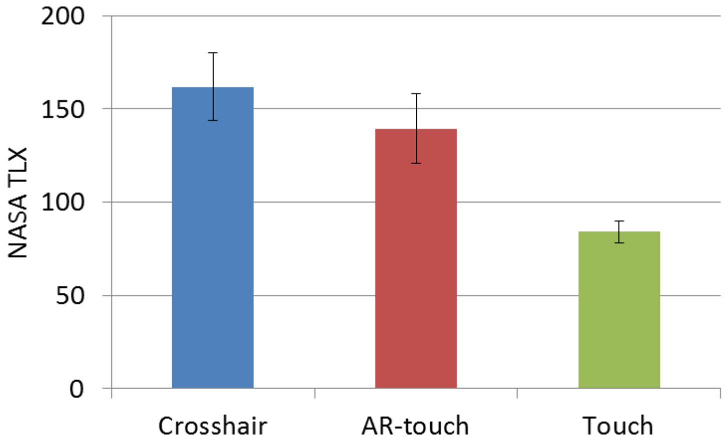


Figure 4.18: NASA TLX for the three interaction techniques to select a region of a photo. Low values mean no task load and high values mean high task load (error bars show standard error).

Another problem participants mentioned is that "my finger is to fat" or with other words "there is the fat thumb again".

4.3.4 Discussion

The results of the first task support our hypothesis that the photo-aligned presentation is more usable than the phone-aligned presentation. Participants perceive the photo-aligned presentation as less demanding and are more satisfied. For the second task the results contradicted our hypothesis. Participants clearly prefer the unaugmented touch technique and the main reason is that image that should be selected does not move.

Based on the sparse comments we assume that the photo-aligned presentation is superior because it provides all information simultaneously and therefore helps to get an overview. The user does not has to select an object to get information about it. This is, however, also the main limitation: The photo-aligned presentation technique does not only allow the user to zoom in and out but the user needs to do so. On pages with a high density of annotations the amount of text that is hardly readable can be confusing. We assume the results can be transferred to other tasks with a similar or lower object density. For those tasks the text size could be further increased, which makes it even easier to get an overview. For tasks with a considerably higher object density the text size must be adjusted accordingly to avoid overlapping texts. In this case an object aligned presentation will presumably become less usable because the user has to "zoom" often by moving the phone towards the objects.

The participants clearly preferred to select regions in a static image compared to the two techniques that use AR. This result is surprising because the design of the study favoured the two other conditions. No zoom was available even though a number of well-

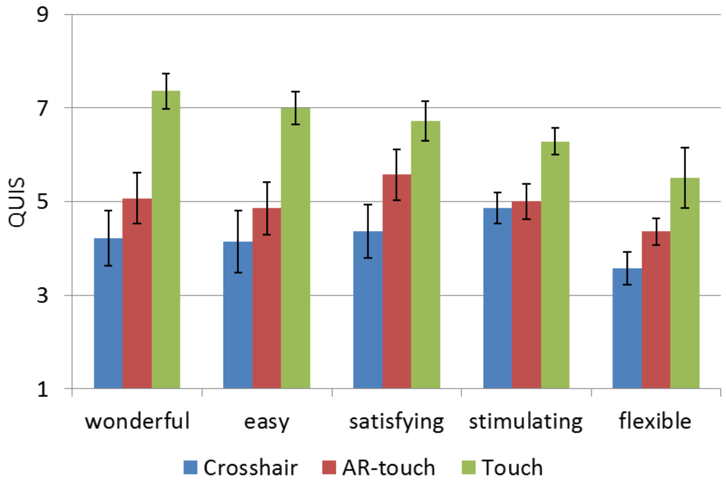


Figure 4.19: QUIS "overall reactions to the software" part for the three interaction techniques to select a region of a photo. High values mean positive reactions and low values mean negative reactions (error bars show standard error).

established techniques exist to implement zooming for static images. The qualitative feedback is also quite clear. Participants prefer unaugmented touch because they do not have to deal with the augmentation. Furthermore, we randomize the order of the conditions but we did not randomize the order of the two tasks. Using handheld AR in the first task certainly improved the participants' performance for the two AR-based techniques due to learning effects. However, as participants had previous experience in using touch screen interfaces they were also well trained in using these interfaces.

The study has two main limitations. The tasks and the setting are artificial in particular for the first task. For the intended use case it cannot be expected that users will search for particular information. Rather, users usually do not have temporal pressure or want to answer specific questions while browsing through a printed photo book. The second limitation, which applies for both tasks, is the short time participants used the conditions. From this perspective, it is even remarkable that all participants could use the conditions of the first task without any problems. Especially for selecting regions more training would certainly improve the performance with the AR-based techniques. However, it is questionable if training can invert the results. Furthermore, users might not be willing to learn using the crosshair-based technique because it is hard to use at least in the beginning.

4.4 Evaluation of the Design Alternatives for Music CDs

Three different interface designs for the augmentation of music CDs have been derived from the participatory study. In order to compare these interface designs, we conducted a user study that is described in the following. In the controlled experiment participants performed two tasks using a within-subject design with the interface design as the independent variable.

4.4.1 Method of the Study

In the controlled experiment participants performed one task to investigate the suitability of the interface designs for information presentation and another task to investigate their suitability for controlling music playback. We assessed the suitability using the "overall reaction" part of the QUIS and the NASA TLX as quantitative measures as well as collecting qualitative feedback.

4.4.1.1 Design

The study is a experiment with a within-subject design that consists of two tasks. In both tasks the independent variable is the interface design consisting of three conditions. The first condition is the object-aligned interface, the second condition is the phone-aligned interface, and the third condition is the two-stage approach. The order of the conditions was counterbalanced to reduce sequence effects.

4.4.1.2 Participants and Apparatus

We conducted the user study with 14 participants, 5 female and 9 male, aged 22-56 (M=29.57, SD=9.53). Half of the subjects had a technical background (mostly undergraduate students). None of them was familiar with the used application or participated in one of the previous studies.

Figure 4.20: Set of music CD covers printed on cardboard used for the evaluation of interfaces to augment music CDs.

A prototype running on a Google Nexus One was used for both tasks. The investigator selected the interface design between the tasks. The interface for the object-aligned condition is shown in Figure 4.21 and the interface for the phone-aligned condition is shown in Figure 4.22. The interface for the two-step condition consists of two views. The view that is used to select objects is the same as for the user study that compares camera-based interaction techniques described in Section 3.5 that is shown in Figure 4.23. For both tasks we prepared three sets of music CD covers printed on cardboard (see Figure 4.20) and annotated each of them with information describing the CD. For the second task we stored the CDs' music tracks on the phone. We prepared an additional set of music CDs for the introduction.

4.4.1.3 Procedure

After welcoming a participant we explained the purpose and the procedure of the study. Furthermore, we asked for their age and noted down the participant's gender. Prior to the first task we demonstrate how to use all conditions.

In the first task, participants had to answer three questions related to the information about the provided CD cover printed on cardboard. To answer each question they had to read the augmentation shown on a mobile phone. E.g. one question was "What is the title of the cheapest CD?". For this example participants must identify the cheapest CD among all CDs and read the annotation that contains the CD's title. After answering a question participants were asked the next question. After completing all questions with one interface design they repeated the task with the next design and another set of CDs. We asked the participants to answer the questions as fast as possible. The information provided by the phone was required to answer a question.

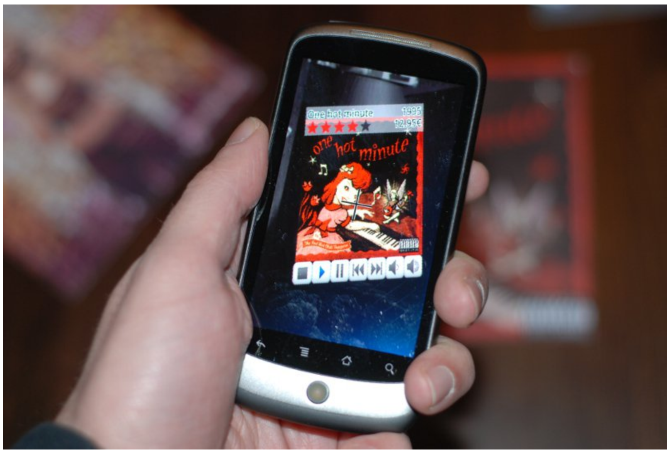


Figure 4.21: The object-aligned interface that aligns the information and playback controls to the CD's top and bottom. The object is highlighted by displaying a coloured image of the CD on top of a greyed out background.

The order of the conditions was counterbalanced. We asked the participants to fill the "overall reactions to the software" part of the Questionnaire for User Interaction Satisfaction (QUIS) [CDN88] to estimate the perceived satisfaction and the NASA TLX

[HS88] to assess their subjective task load.

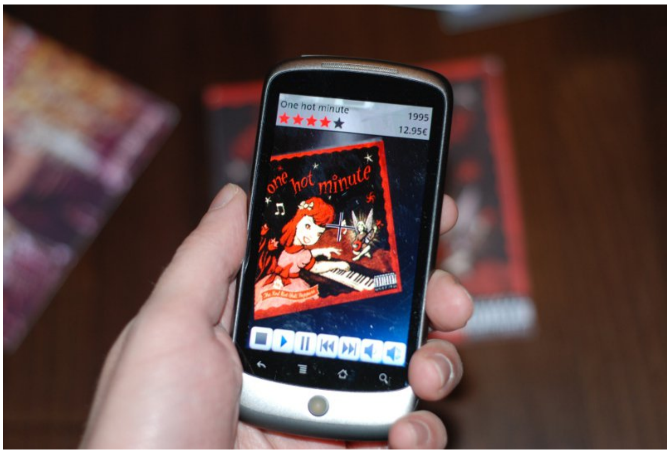


Figure 4.22: The phone-aligned interface that aligns the information and playback controls to the screen's top and bottom. The object is highlighted by displaying a coloured image of the CD on top of a greyed out background.

In the second task, we asked participants to play tracks of the CDs. For each interface designs we asked them to execute tasks such as start the third track of Metallica's Master of Puppets CD. All tasks were formulated in a way that makes it very simple to find the correct CD. Participants were asked to execute the task as fast as possible. After completing the task with one interface design they repeated the task with the next design and a new set of CDs.

The three conditions were counterbalanced to reduce sequence effects. We asked the participants to fill the "overall reaction" part of the QUIS and the NASA TLX after completing the task. As we consider the task completion time as not important for this kind of task and we did not expect relevant differences between the conditions the task completion time is not considered for both tasks.

4.4.2 Hypothesis

For the first task we predicted that the CD-aligned augmentation is more usable than the phone-aligned interface or the two stage approach. With the CD-aligned presentation the user can see all information instantly and can quickly focus on different CDs. Therefore, we assumed that participants perceive this condition as less demanding and give it a lower NASA TLX score. Due to the same reasons we assumed that participants would give a higher QUIS score to the CD-aligned presentation.

For the second task we assumed that the phone-aligned interface would receive a higher QUIS score and that this condition is perceived as less demanding, which would result in a lower NASA TLX score. We assumed that because, compared to the CD-aligned controls, the buttons remain at the same position even if the participant moves

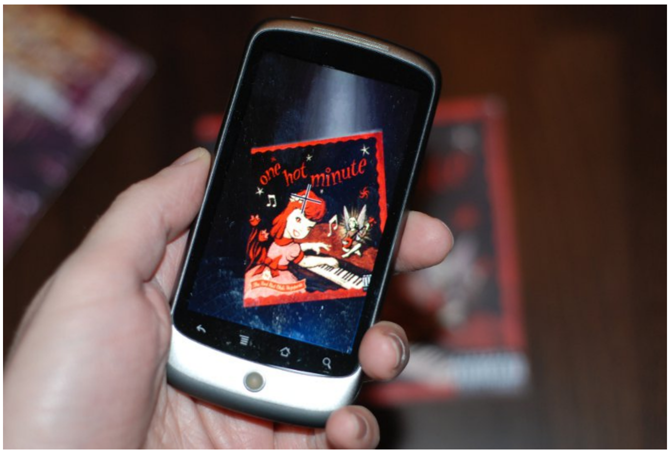


Figure 4.23: The two-step interface that shows no information or playback controls on the screen. The object is highlighted by displaying a coloured image of the CD on top of a greyed out background. The user has to select the CD to go to another view that provides information and playback controls.

the phone. For the other designs we assumed that the CD-aligned interface would be more usable because the user does not need to switch to a second view.

4.4.3 Results

After conducting the experiment we collected and analysed the data. We found significant differences between the three interface designs for the first task as well as for the second task. As the independent variable has three levels we used the Bonferroni correction to reduce the significance levels (i.e. a significance level of $0.5/3 = .01\overline{66}$). Participants' qualitative feedback was translated to English.

4.4.3.1 Finding CDs

Comparing the interface designs an ANOVA shows that the design had a significant effect ($p < .001$) on the QUIS's "overall reactions to the software" part. On average the rating is higher if the augmentation is aligned to the object ($M = 6.84, SD = 1.36$) compared to the score for the phone-aligned design ($M = 5.39, SD = 1.60, p < .01$) and the two stage approach ($M = 4.66, SD = 1.35, p < .01$). The difference between the phone-aligned interface and the two-stage approach is not significant ($p = .09$). The individual scores are shown in Figure 4.24). An ANOVA shows that the interface design had no significant effect ($p = 0.31$) on the NASA TLX score shown in Figure 4.25. The individual scores for the three conditions are as follows: For the object aligned interface $M = 43.64$ ($SD = 21.40$), for the phone aligned interface $M = 43.42$ ($SD = 20.83$), and for the two-step approach $M = 51.14$ ($SD = 18.71$).

We collected few qualitative results related to the conditions. Four of the partici-

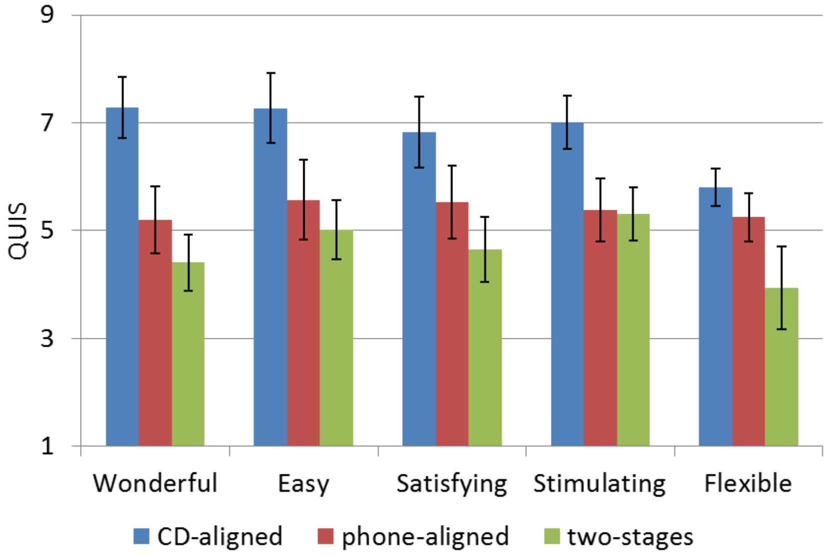


Figure 4.24: Individual QUIS scores of the three interface designs to augment music CDs (error bars show standard error).

pants' comments addressed the performance and the accuracy of the object recognition in general. E.g. one participants mentioned that "the augmentation should be faster" and another participant sad that "it's moving to slow, I cannot go that fast". We observed for both conditions that present the information using an augmentation that some participants prefer to hold the phone sideways. Two participants provided according negative comments about the phone-aligned interface. E.g. the "title should be the other way around" or the system should "rotate the text". About the CD-aligned condition one participant mentioned that "it is faster to switch" and "you can see everything". However, another participant mentioned that this interface requires to "go near the object".

4.4.3.2 Selecting Tracks

In the second task we investigated how well suited the interface designs are to control music playback of the CDs. An ANOVA shows that the interface design had a significant effect on the average QUIS score ($p < .001$). Using the object aligned interface leads to a lower score ($M = 4.26$) compared to the phone-aligned interface ($M = 6.36, p < .01$) or the two-step approach ($M = 6.90, p < .001$). The score for the two-step interface is higher than the score of the phone-aligned interface (see also the individual scores in Figure 4.26) but without a significant effect ($p = 0.46$).

The ANOVA shows that the selection technique also had a significant effect on the NASA TLX score ($p < .05$). Comparing the condition (see Figure 4.27) shows that the CD-aligned interface ($M = 71.36, SD = 10.99$) results in a higher score than the phone-aligned interface ($M = 59.86, SD = 17.08, p < .01$) or the two-step approach ($M = 58.36, SD = 15.13, p < .01$). The score for the phone-aligned interface is lower

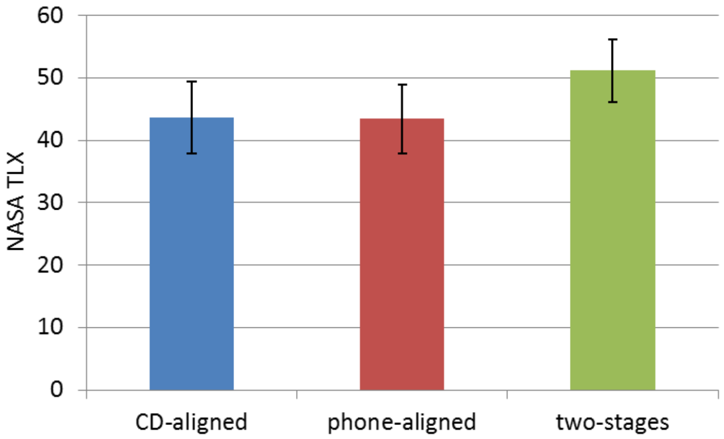


Figure 4.25: NASA TLX scores of the three interface designs to augment music CDs. Low values mean no task load and high values mean high task load (error bars show standard error).

than for the two-stage approach but the effect is not significant ($p = 0.55$).

Participants especially complained about the CD-aligned interface. E.g. one participant stated that "it is difficult to hit the moving buttons" and another one sad "the CD moves whenever trying to start playing" and "it is difficult to catch the button". Participants further noted that "touching changes the position of the phone" but also sad that it is "nice to go quickly from one [CD] to the next". About the phone-aligned interface participants stated that "it is faster than switching to the other screen" compared to the two-stage approach. Other participants, however, said "it is a bit tedious to stay above the CD all the time" or "this is boring".

4.4.4 Discussion

The results of the first task support our hypothesis that the CD-aligned interface is preferred compared to the other interface designs for retrieving information if no further interaction is required. While the participants are more satisfied the interface design had no significant effect on the NASA TLX. Based on the participants' comments we assume the precision and the latency of the algorithms are crucial for the success of a handheld AR applications. Participants criticized both aspects but we assume that better algorithms would further support the condition that is already preferred. Participants like the CD-aligned interface because they can quickly explore a large number of objects by moving from one CD to the other. If using a phone-aligned interface one might consider designing the UI with a landscape orientation as most participants used the prototype sideways

For the second task the results support our hypothesis that the CD-aligned interface is not preferred by the participants and is also perceived as more demanding. Participants

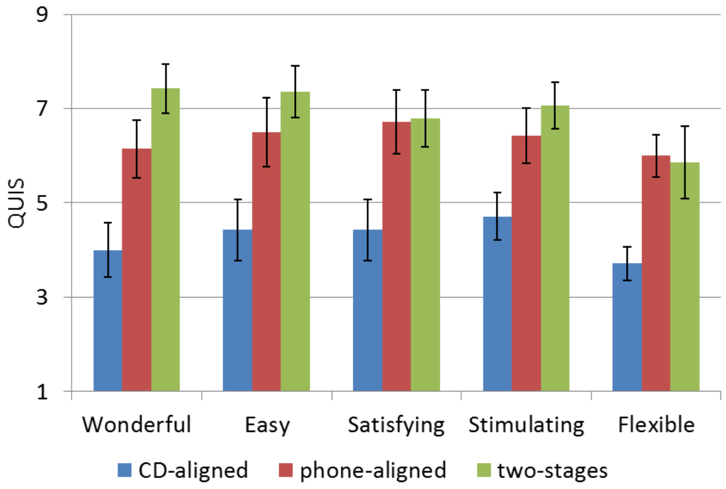


Figure 4.26: Individual QUIS scores of the three techniques to select tracks of music CDs. High values mean positive reactions and low values mean negative reactions (error bars show standard error).

clearly dislike the CD-aligned interface because it is difficult to hit the UI's buttons while simultaneously aiming at the CD. The results do not support our hypothesis that the phone-aligned interface is preferred or less demanding. In fact, the average ratings indicate that the hypothesis that the phone-aligned interface is preferred compared to the two-stage approach might be wrong. This could be the case at least for some participants as we received mixed feedback. Which design is better might depend on the actual use case. If the user wants to quickly select different objects the phone aligned interface is the better choice because the user does not need to switch the views. If the user wants to repeatedly select UI elements, however, a separate view optimized for this task is likely the better option.

The study's main limitation is the lack of training. While we demonstrated the interfaces before starting the tasks the participants used the different variants only for a short time. Trained user might change their preferences after a while. We, however, assume that initial preferences are very important for this kind of application. Users decide if they want to use a mobile "App" in a short time and one cannot expect that they are willing to train. The results of the second task are directly influenced by the size of the provided buttons. The size of the physical object constrains the size that is available for virtual buttons. Using seven buttons distributed over the width of a CD's case results in a fairly large size that is equivalent to 2cm in the physical domain. The button density used in the study was low and we assume that most other use cases would require a higher density. In such cases the advantage for a phone-aligned will be even larger.

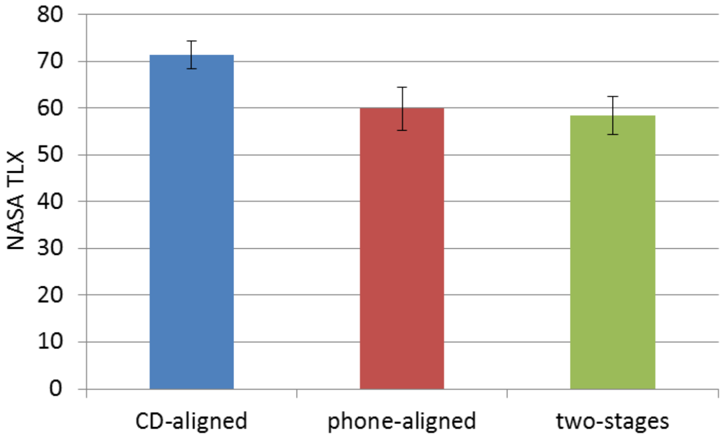


Figure 4.27: NASA TLX scores of the three techniques to select tracks of music CDs. Low values mean no task load and high values mean high task load (error bars show standard error).

4.5 Summary and Implications

After we showed in chapter 3 that interaction with physical media using handheld AR is preferable we investigated the design of handheld AR interfaces. In the following we summarize the conducted studies regarding the interface design of these applications. Furthermore, we outline the implications of our findings on the design of prospective handheld AR applications.

4.5.1 Summary

In this chapter we investigated the design of handheld AR user interfaces for physical media. Two concrete use cases are selected in order to develop potential user interfaces. Printed photo books and music CDs are used as exemplary types of physical media. The same design process is used for both interfaces. Participatory studies are conducted to explore potential interface designs. We collected information and functions from the participants that should be supported by the applications. In addition, we asked participants to sketch interface designs for a handheld AR application using pen & paper. In total 23 subjects participated in the two studies and we collected 40 sketches for the interface designs.

We revised the collected design proposals and identified patterns for aligning interface components. For both use cases participants aligned the interface components either to the phone's border where they remain at a fixed position while moving the phone or they aligned the interface components to the augmented objects and the objects are positioned according to the location of the physical object inside the camera image. Participants also created designs that mix both approaches and proposed a two-step approach where the object must be explicitly selected before using it. In addition, participants suggested highlighting augmented objects by greying out the background or framing the object. For the interaction with printed photo books participants requested the possibility to annotate regions of photos which requires selecting the region first.

Based on the results we designed interfaces for augmenting printed photo books and music CDs. For both use cases we developed an interface that aligns the interface components to the object and another interface that aligns the interface components to the phone's display. An additional interface that requires selecting the object first has been designed for music CDs. As the annotation of regions in photos has been requested by participants, we also designed three techniques to select regions of a photo, two of them using handheld AR. Users can select regions by changing the position and the orientation of the phone relative to an augmented photo, by creating a rectangle with the finger using the augmented photo, or by creating a rectangle with the finger on a static image. Prototypes for all interface designs have been developed for Android smartphones.

To compare the developed designs a controlled experiment has been conducted for each type of media. Altogether, 28 participants took part in the controlled experiments. It is shown for printed photo books and music CDs that an object-aligned interface is better suited to visualize information. E.g. the title of a photo should be displayed relative to the photo and follow it while the photo moves inside the phone's camera image. The most important advantage of this presentation technique is that information about all objects inside the camera image can be presented simultaneously. This enables to quickly explore the information provided by different objects. In contrast, we found that tapping the augmentation, which is required if input controls are aligned to the object is difficult. While the user tries to hit a button with the finger the phone must kept still at the same time. This proved to be too laborious and the advantage of being able to access multiple objects simultaneously cannot compensate this disadvantage. For selecting regions of physical objects to annotate object in photos we found that users prefer to select regions on static images compared to selection using AR.

4.5.2 Implications

We designed, implemented, and evaluated the user interface of handheld AR applications for two different use cases. This enables to assume that the results can be transferred to other application domains beyond the particular types of physical objects used throughout the studies.

We found that information connected to a physical object should be displayed in the reference system of the object. Users can get a quick overview and explore available content in a self-directed way. This finding can easily be transferred to other types of media such as DVDs, Books, or product packages if a quick overview is desired. For the augmentation of paper maps the POIs themselves but also their textual description should be presented aligned to the map and not to the phone's display. Developing paper-based handheld AR applications in general the findings should be considered when designing the user interface. Looking at previous work we find a number of examples that do not conform to this finding. E.g. the marker-based handheld AR demonstrator Rohs developed in his early work for interaction with paper forms presents information aligned to the phone's display [Roh05]. Similarly the prototype Toye et al. [TSM+07] used to study the interaction with mobile services using marker-based handheld AR presents information above the camera image at a fixed position. Another example that augments transit maps [SMGB11] uses a mixed design and displays textual information at the top of the display. While these prototypes have not been intended for productive use it can be assumed that their usability could be clearly improved using an object-aligned interface.

While information should be presented aligned with the object we found that input controls that must be touched with the finger should not. Pressing simple buttons to control music playback turned out to be difficult and is perceived as demanding if the buttons are displayed in the reference system of the object. The results for selecting regions of printed photos also show that touching in the reference system of the object is challenging. Users have to focus on touching the touch screen and have to hold the phone in a fixed position at the same time. We conclude that input controls should always be aligned with the screen and remain in the reference system of the phone. This finding is relevant for all physical objects that should provide some interactivity. E.g. DVDs, newspaper, or textbooks could provide multimedia content that can be controlled.

Even though the physical objects we used can be easily distinguished from the background participants proposed to explicitly highlighting augmented objects. It has been proposed to gray out the background and to frame the object with a border. In fact, different ways of highlighting are needed for certain applications. Highlighting the physical object itself has been demanded in order to make them identifiable. Thereby the user is also ensured the system successfully recognizes the object. Subjective feedback from participants showed that greying out the background and displaying only the objects with colours was well received. This approach can be transferred to use cases where small and medium objects, such as CDs, photos, or physical media in general, are augmented. It can also be necessary to highlight certain regions of an object such as persons in printed photos. Drawing a rectangle around the region is a simple way that avoids ambiguity with the highlighting technique used for the physical objects itself. Highlighting regions of an object is necessary for certain types of media, such as paper maps, text documents, and other objects that are composed of different fragments. We assume that our highlighting approach can successfully be applied to these types of media.

All interfaces are designed to be used while holding the phone upright. Still, we observed that some participants held the phone sideways. Participants invested the additional effort of mentally rotating the presented text to be able to keep the phone in a sideway orientation. Some participants even commented that the orientation of presented text is wrong. An advantage of holding the phone sideways is that the phone can be held with two hands. Thereby, it is easier to keep the phone still which reduces jitter in the augmentation. The results suggest that the interface of future handheld AR applications should be in landscape or provide the option to switch to a landscape mode.

From the users' perspective the used algorithms still leave room for improvement.

Participants criticized the precision and the speed of the object recognition and object tracking algorithms during both studies and all interface designs. It can be concluded that effort invested in improving the algorithms is actually worth the hassle. Even though, the phone-aligned interfaces are less affected by jitter and lag, the object-aligned interfaces still turned out to be more usable. Compared to the phone-aligned interfaces an Object-aligned interface might require an additional estimation of the object's pose. Therefore the algorithm for pose estimation is actually required and research about lightweight pose estimation is well-invested.

5 Off-Screen Visualization in Handheld Augmented Reality

In order to interact with physical objects using any camera-based interaction technique the user must know that augmentable physical objects exists nearby. Furthermore, the user must determine the location of these objects. For some types of physical objects, such as photo books and music CDs used in the previous chapter, this is a trivial problem for the user. For other types of objects determine their availability and location can be challenging and even frustrating (see 3.4). The nature of points of interest (POIs), for example, can be diverse including monuments, buildings, squares, and even plants. Therefore, applications that offer a camera-based interaction, in particular handheld AR, must support the user in determining the existence and position of nearby objects that can be augmented by the application. So called off-screen visualization techniques consider mobile devices' screens as a window in a larger space. Such visualizations point at objects that are located outside of this window. They have successfully been used for computer games (see [BR03]), digital documents [ZMG$^+$03], and digital maps on mobile devices [BR03, BCG06, BC07, GBGI08].

In order investigate how off-screen visualizations can be applied to handheld AR we investigated three different aspects by conducting five consecutive user studies. Since we are not aware of earlier studies that investigated off-screen visualizations for digital maps using static maps and accordingly quite artificial tasks that did not involve panning the map [BR03, BCG06, BC07, GBGI08] we reinvestigated the visualization of off-screen objects for digital maps. The aim of the initial two studies is to find a starting point for developing off-screen visualizations for handheld AR. Their contribution is twofold: First we confirm that arrow-based visualization outperform the circular approach Halo and second we show that it is possible to conduct controlled experiments[1] by publishing applications in publicly available mobile application store. In the subsequent study we analyze the effect of off-screen visualizations on the interaction with augmented physical maps. We show that an arrow-based off-screen visualization significantly improves the user's performance. Finally, we investigate off-screen visualization for 3D objects by developing and comparing three visualization techniques for highlighting POIs. Based on the results we revised the design of the visualization and compare it in a controlled experiment with a baseline. We show that the developed off-screen visualization outperform the commonly used mini-map for handheld AR applications.

5.1 Off-Screen Visualizations for Digital Maps

Visualizing off-screen objects has received some attention for interaction with digital maps on mobile devices. Zellweger et al. [ZMG$^+$03] introduced City Lights, a principle for visualizing off-screen objects for hypertext. An extension of the City Lights concept

[1] Controlled in the sense that we have control about assigning conditions to participants to randomly assign them and that the conditions are exactly the same apart from the respective visualization technique.

for digital maps is Halo [BR03]. For Halo circles that intersect the visible area shown on the device's display are drawn around the object. Users can interpret the position of the POI by extrapolating the circular arc. Baudisch et al. showed that participants are faster when using Halo and prefer it compared to arrows with a labelled distance but make more errors [BR03]. Burigat et al. [BCG06, BC07] reviewed these results by comparing Halo with different arrow types e.g. by visualizing distance through scaling the arrows. They found that arrow-based visualizations outperform Halo, in particular, for complex tasks. Other off-screen visualization have been developed (e.g. Wedge [GBGI08]) but we are not aware of studies that clearly show that these outperform existing approaches. In general, the previous work conducted studies with static maps that participants had to interpret. E.g. they did not consider tasks where users can dynamically interact with the map by panning it. Furthermore, the results of previous studies are not consistent and the conclusions are based on the performance of less than 17 participants which share similar backgrounds (e.g. computer scientists).

Previous work could not clearly show which visualization technique is best. To base our work on applying off-screen visualizations to handheld AR we conducted two studies to determine the performance of different off-screen visualization techniques for digital maps. We compare the three visualization techniques Halo, stretched arrows, and scaled arrows shown in Figure 5.1. To provide a definite answer to the question, which technique is preferable we aimed at attracting a large number of participants with various backgrounds that execute controlled tasks in their natural environment. Furthermore, we aimed at a task that is more realistic than tasks used in previous work by requiring panning of the map.

With the introduction of mobile application stores such as Apple's App Store and Google's Android Market a new way to conduct user studies became available to the average HCI researcher. The Android Market, in particular, enables to publish an application in a few minutes without any review process. By publishing applications in mobile application stores, researchers benefit from a worldwide audience. They gain access to participants with various cultural backgrounds and different contexts [MMB$^+$10]. By developing "Apps" with the aim to answer specific research questions and logging the user's behaviour it is possible to harvest a large amount of data samples. In the following we report about two studies that compare three off-screen visualizations using an app published to the Android Market as an apparatus. To our knowledge these studies are the first showing that it is possible to conduct controlled experiments via mobile application stores.

5.1.1 Comparing Off-Screen Visualizations with a Tutorial

To compare the three off-screen visualizations studied by Burigat et al. [BCG06, BC07] including panning of the map we implemented an application for the Android platform that is called Map Explorer. We considered two main requirements for the application: It must allow the comparison of the three visualizations techniques and it must at least

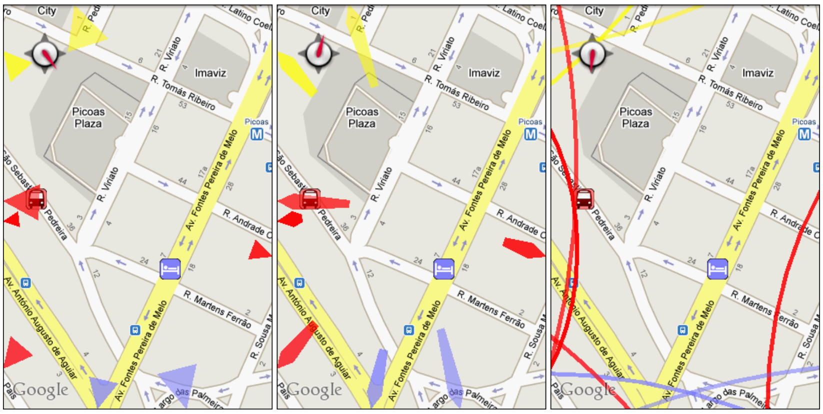


Figure 5.1: The three off-screen visualization techniques for digital maps: scaled arrows, stretched arrows, and Halos.

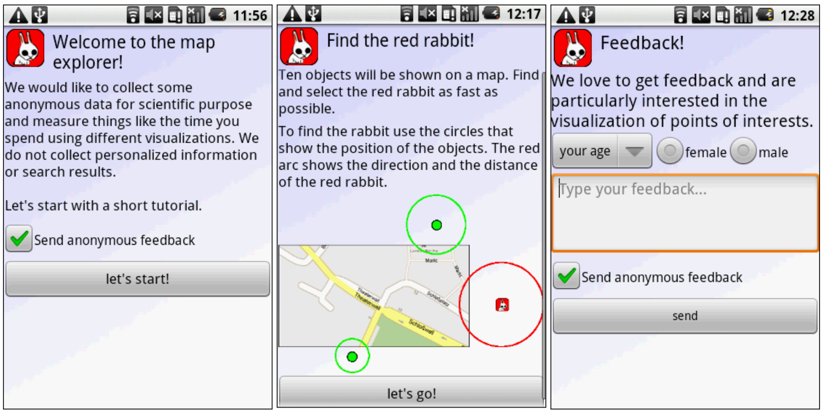
pretend to be useful to attract a sufficient number of participants. To make the application somewhat useful we developed a location-based application that presented POIs on a map. The visualization techniques are compared using a tutorial that requires executing the same task using each technique.

5.1.1.1 Developed Prototype

A number of location-based applications that try to attract potential users' attention are available in the Android Market. This includes the commercial Google Maps that is pre-installed on Android devices but also research prototypes such as the PocketNavigator [PPB10]. As these apps compete for potential users' attention this application domain is very competitive. Thus, we aimed at an approach that collects meaningful data even from participants that use the app for a short time.

To attract an adequate number of users the application must be downloaded and installed at own will. From our previous experience with applications in the Android Market we assumed that an important factor is the user ratings of the application. Thus we intended to make the application useable and useful enough to be accepted by a sufficient number of users. A map that is based on Google Maps is presented in full-screen. The application offers the standard functionalities of a location-based application. Users can search for nearby POIs and access details about the POIs including reviews, ratings, and images using the web services of either "Qype" or "Yahoo! local" as the data source. While the application is used the time using each off-screen visualization is measured. We also measure if the user interacts with the application or not. Furthermore, users can fill the feedback form shown in Figure 5.2.

We implemented the three off-screen visualizations Halo, stretched arrows, and scaled arrows shown in Figure 5.1. To make the visualizations comparable we decided for a tutorial which mimics the well-defined tasks usually found in lab experiments. Using defined tasks should improve the repeatability and reduce the effect of other influences.



Welcome to the map explorer!

We would like to collect some anonymous data for scientific purpose and measure things like the time you spend using different visualizations. We do not collect personalized information or search results.

Let's start with a short tutorial.

☑ Send anonymous feedback

let's start!

Find the red rabbit!

Ten objects will be shown on a map. Find and select the red rabbit as fast as possible.

To find the rabbit use the circles that show the position of the objects. The red arc shows the direction and the distance of the red rabbit.

let's go!

Feedback!

We love to get feedback and are particularly interested in the visualization of points of interests.

your age ▾ ○ female ○ male

Type your feedback...

☑ Send anonymous feedback

send

Figure 5.2: Tutorial instructions and feedback form of the prototype to compare different off-screen visualizations for digital maps.

The tutorial starts with an introductory text and consist of a simple find-an-select task for each visualization afterwards (see Figure 5.2). Users are informed that we collect data for scientific purpose while they use the application. By deselecting a check box users can opt-out and we accordingly collect no data from them. Starting our app a tutorial is shown that asked the user to execute the same task using each of the three visualization techniques. Participants were free to quit the tutorial and the application at any time.

5.1.1.2 Method of the Study

In the controlled experiment participants performed a single task to compare the three visualization techniques. A within-subject design with one independent variable resulting in three conditions was used. We had no control over the number of participants, their age, or their gender. Nonetheless, we aimed at attracting some thousand users to collect a sufficiently large sample. As participants installed and use the app on their own phone we expected diverse devices. Because this study is, to our knowledge, the first controlled experiment distributed via a mobile application store, little experience existed about the participants one could expect. While it cannot be expected to attract participants that represent the perfect sample of the global population it can be assumed that participants are a good sample of the population of smartphone users. In [HPP$^+$11] we provide an overview about the participants one can expect that is based on a number of studies that use apps distributed via mobile application stores as an apparatus.

The procedure of the experiment was controlled by the tutorial that is shown when the app ist started. From a user's perspective the aim of the tutorial is to explain the visualization techniques while our aim was to measure the users' performance while completing the tasks provided in the tutorial. While executing the task of the tutorial a map is shown in full screen. The map contains 10 POIs that are randomly distributed around the user's position. The maximal distance of the POIs from the centre of the display is 2.5 times the height and width of the screen. One POI is the target (the only red object) that should be selected. This POI is not initially visible on the screen to ensure that it "off-screen" and all users must pan the map at least once. The map can be explored by panning it with the finger just as the standard Google Maps on Android. We did not allow zooming while using the tutorial in order to have more control and to ensure that panning is necessary to complet it. POIs are selected by tapping on its icons. The tutorial is automatically started if the application is started for the first time. To reduce sequence effects the order of the visualization techniques is randomized. We measure the time required to complete each of the tutorial steps and we determine the number of map shifts by counting how often a user pans the map. The collected data is send to our server every 30 seconds and after a user finishes the tutorial.

We expected that our results will be consistent with the results described by Burigat et al. [BCG06, BC07] and hypothesized that users will be slower and need to pan the map more often with Halo. Because of their similarity we expected only negligible differences between the two types of arrows.

5.1.1.3 Results and Discussion

The Map Explorer was published in the Android Market on the 2nd of April 2010[2]. We did not actively advertise the app among our friends and colleagues. In the following we report the results derived from the data collected until the 26th of May 2011. According to the statistics provided by Google's Android Developer Console the app has been installed 8035 times. In total we collected samples from 6053 accounts. 1098 participants from 67 different locales and 57 different Android devices completed the tutorial.

The average number of map shifts and the task completion time is shown in Figure 5.3. An ANOVA shows that the visualization technique had a significant effect on the number of map shifts (p<10^{-5}). As we used three conditions the significance levels for the follow-up t-tests are reduced to $0.5/3 = .01\overline{66}$ with a Bonferroni correction. Participants panned significantly more often using Halos (M=10.17, SD=13.15) than using stretched arrows (M=8.25, SD=11.10, p<.0001) or using scaled arrows (M=7.77, SD=11.09, p<.0001). The difference between stretched arrows and scaled arrows is not significant (p=0.15). Consequently, participants also spend significantly more time using Halos (M=16.18s, SD=19.94s) than using stretched arrows (M=13.68s, SD=19.09s, p<0.01) or using scaled arrows (M=13.25s, SD=19.09s, p<.001). Again the difference between stretched arrows and scaled arrows is not significant (p=0.30).

Using Halo participants are slower and need more map shifts than using an arrow-based visualization. The differences between the two arrow-based visualizations were very small for both depended variables and we did not find a significant effect. We did not receive useful comment with the feedback form.

The obtained results support our hypothesis. However, the results seem to contradict earlier work that compared Halo with arrows that are labelled with the distance for

[2] An updated version of the Map Explorer is available in the Android market: `https://market.android.com/details?id=de.offis.map` last accessed 24 November 2011

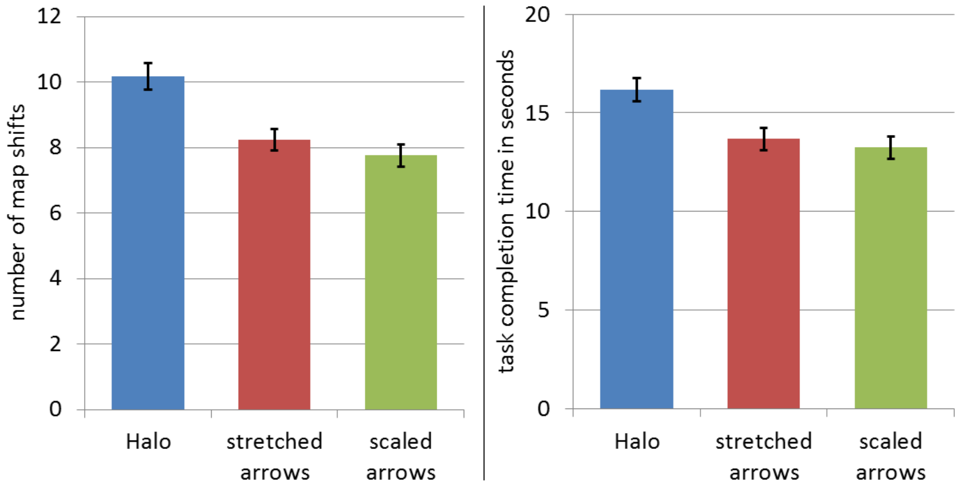


Figure 5.3: Average number of map shifts (left) and the average task completion time of the comparison of the three off-screen visualizations using a tutorial. (right) (error bars show standard error).

static maps. In line with more recent work [BCG06, BC07] we assume that scaling or stretching the arrows to communicate the POIs' distance is much easier to interpret that showing the distance with numbers. Thus, the results are consistent with previous work [BCG06, BC07] but suggest their findings might be generalized to much more realistic tasks that involve panning the map.

However, further investigation of the collected shows that a number of users needed much more time than one would expect (e.g. longest time spend using Halos was 100 seconds). Reconsidering our design it might be assumed that instead of measuring the pure task completion time the results are affected by the "interestingness" of the visualizations. From informal tests we can report that some users explore the map much longer using Halo than using the other visualizations because of Halo's aesthetics. Furthermore, our results are also limited because users had no previous training and performed the tasks only once with each visualization.

5.1.2 Comparing Off-Screen Visualizations with a Game

The results of the comparison described in the previous section suggest that the arrow-based visualizations are preferable compared to the alternative visualization Halo. The method of the study, however, leaves room for interpretation. On the one hand the differences could be caused by Halo's interestingness. Participants might not understand Halo without detailed explanation and want to explore how the visualization changes when they interact with the map without focussing on the tutorial's tasks. On the other hand the results are likely heavily affected by their intuitiveness. Participants received not training and performed the task only once using each condition and we therefore could not observe training effects. Therefore, we conducted another study with similar aims as for the study described in Section 5.1.1. In particular, we aimed at answering the following questions:

- How do different techniques for visualizing off-screen objects perform in an interactive task that involves panning the plane.
- How do the visualization techniques scale if the number of shown objects increases?
- How easy are the visualization techniques to learn and do users understand the meaning of the respective visualization without lengthy instructions?

5.1.2.1 Developed Prototype

Again we aimed at comparing the three visualization techniques by conducting a "controlled" experiment. This leads to the three conditions Halo, stretched arrows and scaled arrows. A repeated measurement design reduces the effect of the individuals compared to an independent measurement design. In a public experiment one cannot control important aspects such as the selection of participants, used devices and the participants' context which is why we decided for a repeated measurement design. In order to investigate the scalability of the visualization techniques multiple tasks with different numbers of objects are used.

It is crucial for public studies to motivate people to participate. Even though the visualizations have been designed for maps it would be difficult to force a mobile user looking for a hotel to repeat the same task with a different visualization technique. Therefore, we decided to use a mobile game which enables to naturally confront participants with variations of the same task. Thereby, it can be assured that participants repeat the same tasks while only the independent variables (i.e. the visualization technique and the number of visualized objects) are varied. However, as the game has to be installed and played by users at their own will it is necessary to find a balance between validity of the study and fun of the game.

We decided to use an increasing level of difficulty to motivate players. A game starts with a stage of three levels each containing 30 objects, represented by "cute" rabbit icons. The objects are randomly distributed on plane that can be paned much like a digital map. Each level uses a different off-screen visualization (see Figure 5.4). The task of the player is to "poke" as many objects as possible by tapping them with the finger in a certain time frame. Once an object is poked it fades to grey and a new object appears. If a player finishes the three levels he or she goes to the next stage where 20 objects are used and afterwards to a stage with 10 objects. The visualizations are randomized within a stage to reduce sequence effect. After finishing three stages the game starts from the beginning with more time to complete a level but also with more objects needed to successfully finish a level.

We implemented the game for the Android platform. The visible area covers the same fraction of the complete field on different devices by scaling a fixed fraction to the whole screen. It is slightly affected by different devices' aspect ratio. A short explanation (see

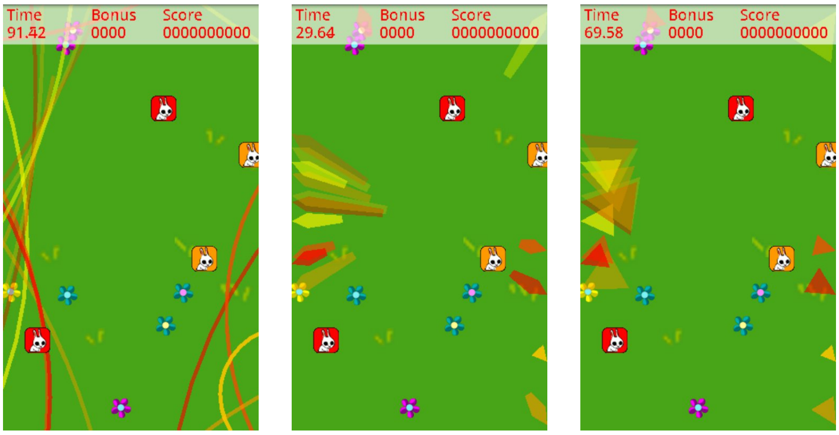
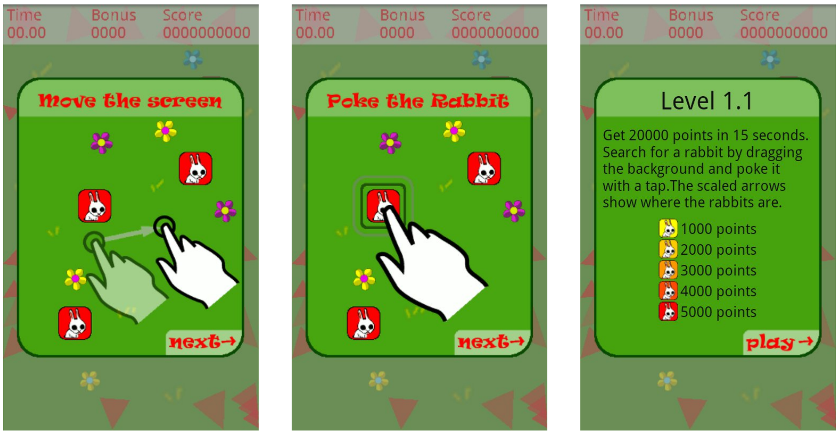


Figure 5.4: In-game screenshots of the three visualization techniques Halos, stretched arrows, and scaled arrows.



Get 20000 points in 15 seconds. Search for a rabbit by dragging the background and poke it with a tap.The scaled arrows show where the rabbits are.

- 1000 points
- 2000 points
- 3000 points
- 4000 points
- 5000 points

Figure 5.5: Screenshots of the introduction for the game that compares the three off-screen visualizations for digital maps.

Figure 5.5) is shown each time a game is started. Furthermore, the player gets scores each time a rabbit (i.e. object) is tapped. A bonus is added if the player taps multiple rabbits in a row. To increase the motivation we implemented a local and a global high score list which can be accessed from the main menu. Furthermore, we added music that is played during the game. Each time a level is finished the number of tapped rabbits and the particular level is transmitted to our server. We also log the device's time zone, the selected locale, the device's type, and an anonymized device id.

5.1.2.2 Method of the Study

In the controlled experiment participants performed a single task to compare the three visualization techniques. A within-subject design with two independent variables was used. The first independent variable is the visualization technique (scaled arrows, stretched arrows, and Halo) and the second independent variable is the number of rabbits (10, 20, and 30). The order of the off-screen visualizations was randomized to reduce sequence effects. The number of rabbits, however, was descending. In the first three levels 30 rabbits were used, followed by three levels with 20 rabbits and three levels with 10 rabbits. As participants were free to play any number of level each of them could contribute any amount of measurement. As the only depended variable we measure the number of targets that a player selects in each level.

As in the previous study we had no control over the number of participants, their age, or their gender. Based on the results of the previous study, we aimed at collecting data from a few thousand participants, expected diverse devices, and players from all over the world. Even though, games are a very particular type of application we assumed that participants are a good sample of the population of smartphone users.

5.1.2.3 Results and Discussion

The describe game was published in the Android Market on the 14th of April 2010[3]. We did not actively advertise the game among our friends and colleagues. In the following we report the results derived from data collected until the 25th of June 2010. According to the statistics provided by Google's Android Developer Console the game has been installed 4371 times. In total we collected samples from 3934 accounts. These samples came from 40 different types of devices. The devices cover most of the diversity of the Android phones available at that time. E.g. the most frequent Sholes (alias Motorola Droid) runs Android 2.1 and has a 3.7" (854x480px) screen while the second most frequent HTC Hero running Android 1.6 has a 3.2" (480x320px) screen. The most frequent locale is en_US with 68.3%. In total English locales accounted for 76.5% and more than 92.3% use a western language. While users can freely select the used locale the results are very consistent with the observed time zones.

We analysed the effect of the visualization technique on the players' performance if different numbers of rabbits are present. Since different levels have different durations we normalized the number of poked rabbits to "hits per minute" (hpm). Furthermore, we pre-processed the raw data by removing incomplete samples and samples where players did not poke a single rabbit. An ANOVA shows that the visualization technique significantly affected the players' performance for 30, 20, and 10 rabbits (all $p < .05$). The average performance is shown in Figure 5.6. With 30 rabbits and using scaled arrows (M=38.41hpm) the players archived a higher performance (both $p < 0.01$) than using Halos (M=37.33hpm) or stretched arrows (M=37.26hpm). When 20 rabbits are used players achieve a lower performance with Halos (M=36.75hpm) than with stretched arrows (M=37.82, $p < .05$) or scaled arrows (M=38.29, $p < .01$). With 10 rabbits the order of the visualizations is reversed. Using Halos (M=35.33) players perform better than using stretched arrows (M=33.52, $p < .001$) or scaled arrows (M=32.18, $p < .001$). The difference between stretched arrows and scaled arrows is also significant ($p < .05$).

[3] An updated version of the game is available in the Android market: `https://market.android.com/details?id=net.nhenze.game.offscreen` last accessed 24 Novermber 2011

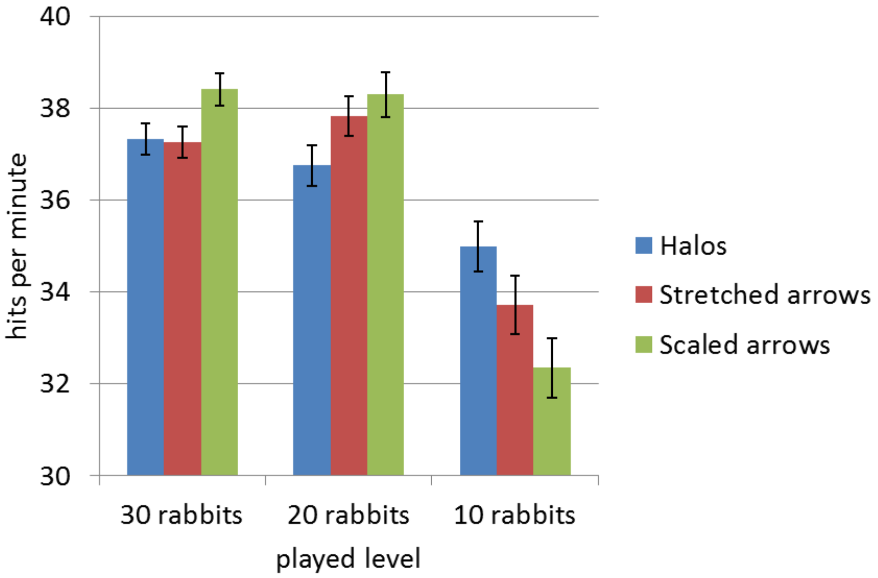


Figure 5.6: Comparison of the three off-screen visualizations for digital maps for different numbers of objects.

We expected that the learning curves for the three visualizations differ. We assumed that the arrow-based visualizations are more intuitive and novice players perform better with them than with Halo. The design of the experiment does not allow a systematic analysis but the players' performance after playing a respective number of levels shown in Figure 5.7 suggest a tendency. The trend lines of the three techniques are very similar and we therefore assume that their learnability is also surprisingly similar.

Due to the nature of the study we could not control which device the participant uses. The large number of different devices (40) makes Type I errors (i.e. we believe that there is an effect, when in fact there is not) very likely if we do a pair wise comparison of all devices. Furthermore, the numbers of samples from the devices are very different and devices with a low number of samples should not be considered. In addition, it is possible that players with a low performance (partly induced by the used device) quit playing the game early which would make the differences between devices look larger than they actually are. As we did not define a procedure beforehand (e.g. how many samples are needed from each device) it is likely that extensive analysis would be error-prone. Therefore, we only exemplarily compared the two most often observed devices. The average hits per minute for the Sholes is 39.37hpm (n=2205) and 34.57hpm for the HTC Hero (n=1134). Even with a very conservative significance level the average number of hits per minute significantly differs ($p < 10^{-9}$)

In summary, the results show that the visualization techniques scale differently. For 30 objects arrows are more suitable and for 10 objects player perform better with Halos. The difference between the visualization techniques regarding learnability is presumably small. As expected, the used device does affect the players' performance.

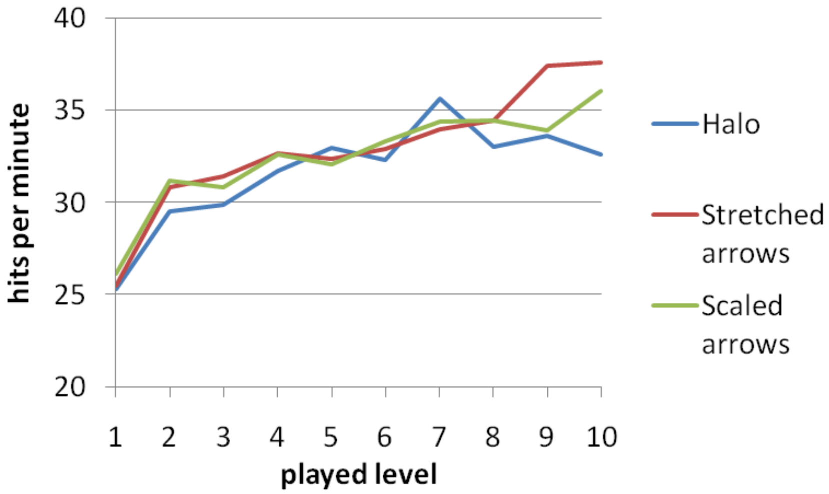


Figure 5.7: Players' average performance with the three off-screen visualizations after playing a particular number of levels. Only samples where players poke at least one rabbit are considered.

For a large number of objects our results are consistent with the results of previous work that used complex tasks, static maps, and a low number of objects [BCG06, BC07]. In contrast, our results suggest that Halos perform better than the arrow-based approaches for a low number of objects. This, is consistent with [BR03] which used a very low number of objects to compare Halos and arrows with labelled distances. However, our study analysed off-screen visualization using a task that requires to dynamically interacting with the objects while in previous studies the participants used static maps. Thus, our results are particularly relevant for interactive systems such as common digital maps like the popular Goggle Maps.

The study treated internal validity for external validity. Due to the large number of participants with different background, devices, and contexts our results are more generalizable than studies involving 12 [BR03] or 17 [BCG06, BC07] participants, which use the same device, perform the tasks in the same room, and live in the same region. Even though we tried to address users from all over the world most players originate from the US or another western country. It might be possible to attract more players from other cultural backgrounds by internationalizing the game and its description in the Android Market. The experiments internal validity is limited because we had little control over external factors and the data is heavily affected by noise. This is one of the reasons why we can conclude little about learnability and differences between devices.

5.1.3 Conclusions

In two studies we compared off-screen visualizations that show the position of objects on a 2D plane. In contrast to previous work [BR03, BCG06, BC07, GBGI08] the task that participants executed required to pan the plane. Both studies support the conclusion that the arrow-based visualization outperform Halo – at least if more than 20 objects are presented simultaneously. The differences between the two arrow-based techniques are, however, small and we could not find a significant effect. We assume that the reason why arrows are superior is that they allow determining an object's direction more easily or more precisely.

In total, we attracted 5,032 participants from all over the world. Therefore, the studies do not only allow strong conclusions about off-screen visualizations but also show that conducting controlled experiments using mobile applications stores is possible. Only few studies (e.g. [ZKG$^+$09, MMB$^+$10]) have been conducted using this distribution channel prior to our work. Furthermore, to our knowledge, we have been the first who showed that distributing apps via mobile application stores can also be used for controlled experiments. Therfore, our work contributes to a new exiting field that enables to conduct mobile HCI studies with a very high external validity and a large number of participants (see e.g. [HPP$^+$11, HRB11, HB11a, PHB11]).

5.2 Off-screen Visualizations for Printed Maps

Map navigation with handheld devices helps mobile users to understand and explore their current location. However, interaction with digital maps is limited by the device's inherent small screen size. It is often difficult to identify and comprehend the distribution and position of landmarks using maps shown on rather small mobile devices. Traditional scrolling and panning interfaces with joystick or touch screen input offer only limited support in exploring large-scale maps on those small displays. Paper maps and public maps are often found in the city centre to provide an overview about an area. However, paper maps only contain generic information and places of general interest. More specific information, such as the locations of ATMs, shops, and restaurants as well as short-lived events are omitted because of the limited space and the static nature of paper maps.

Handheld AR can be used to combine the visualization of detailed and personalized information provided by digital maps with the provision of an overview by a paper map. The idea is to provide an overview through the physical surface while personalized information is displayed on the phone's screen. By determine the phone's position in relation to a static paper map, dynamic information can be merged with the camera video that is presented on the phone's screen. Handheld AR can be used to augment static paper maps with a higher level of detail, personalized information, or short-lived events. An example is to provide tourists with up-to-date information about nearby events, ratings of restaurants, and routes.

A concurrent approach that relies on the mobile phone's physical position and movement is using the device as a dynamic peephole (hereinafter referred to as peephole) [Yee03, MWW06] that serves as a window into a virtual space. Similar to the handheld AR the user moves the handheld device and the visualization is updated according to the

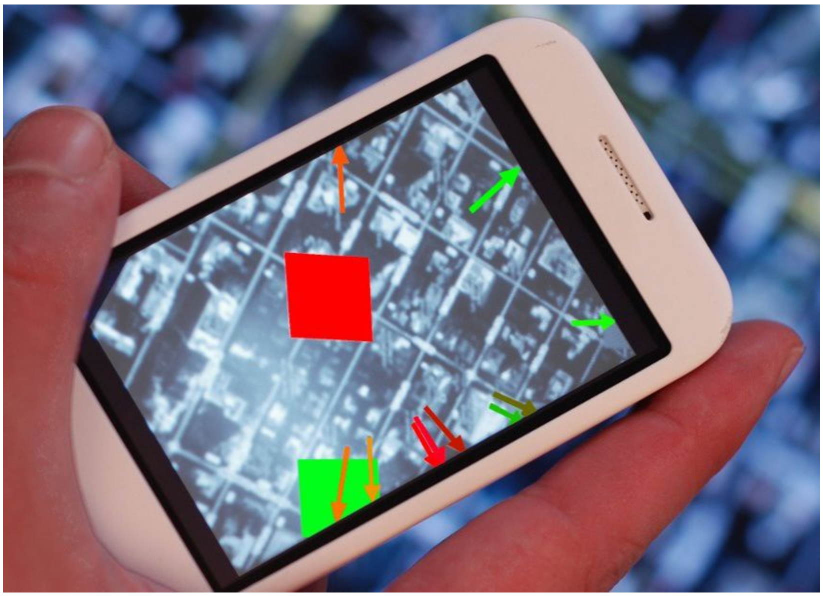


Figure 5.8: Physical map augmented using a mobile phone. POIs on the map are shown with coloured rectangles and arrows pointing at object beyond the display.

device's position. In contrast to handheld AR, information is not visualized relative to a physical surface but relative to a virtual surface. Thus, the handheld AR and the peephole interface provide the same visualization on the mobile device's screen. The difference is that handheld AR provides additional visual context through the underlying physical surface while the peephole interface only visualizes information on the mobile device's screen itself.

Both techniques have been used to show POIs on maps (e.g. [RSR$^+$07, MOP$^+$09]). In this case, the difference between the peephole interface and handheld AR is that the paper map used in conjunction with the AR enables users to locate POIs on the map that are currently not visible on the mobile device's screen. Conceptually the paper map is used to visualize the "off-screen objects". In the following we compare the off-screen visualization provided by handheld AR with off-screen visualizations developed for maps shown on mobile devices. Visualization techniques for off-screen objects are applied to a handheld AR and a peephole interface. We show that using an arrow-based visualization for off-screen objects in combination with handheld AR (as shown in Figure 5.8) or in combination with a peephole interface lowers the task completion time and decreases the perceived task load.

5.2.1 Developed Prototype

In order to study the effect of using an off-screen visualization in combination with handheld AR or a peephole interface we implemented both interaction techniques. A

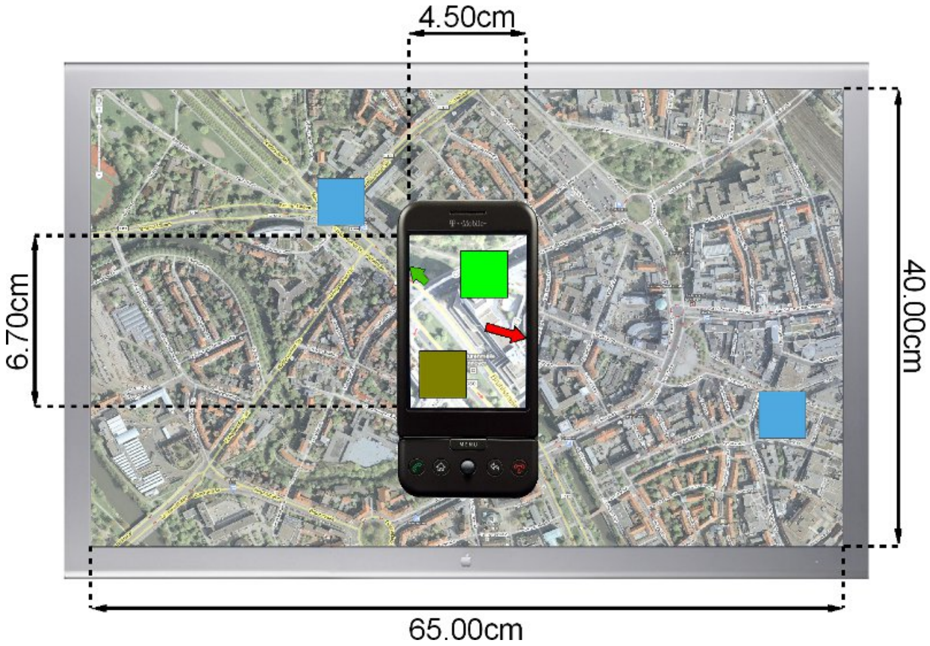


Figure 5.9: Concept of the off-screen visualization for augmenting physical maps and the specification of the external display. The positions of two off-screen targets are shown by the blue rectangles for illustrative purpose.

map which is displayed on a large screen is used as static surface. We used a display instead of a paper map to be able to quickly exchange the map during a user study. In order to make the conditions as similar as possible the screen shows a map in all conditions. I.e. the map itself was always visible but the positions of the POIs are only visible if handheld AR is used. For the peephole interaction the screen showed a bare map without POIs. The map was shown for the peephole condition to eliminate effects on the user caused by different backgrounds.

The location of the POIs was chosen randomly. Using handheld AR the POIs are marked with blue rectangles on the underlying map. A mobile phone was used for both interface. The phone displays the video from its rear camera. If the phone's camera is pointing at the map an augmentation is embedded in the video. The position of the POIs is marked with coloured rectangles on the phones screen. As we showed in Section 5.1 that arrow-based off-screen visualization outperform the concurrent approach Halo, stretched arrows are used to visualize POIs which are currently not visible in the camera image. The arrow's length is scaled according to the distance of the respective POI from the edge of the phone's screen. Each arrow has the same colour as the respective POI. Figure 5.9 shows the visualization with arrows pointing towards off-screen objects. In this example the object's position is also displayed on the underlying map.

We implemented the prototype using a 30" Apple Cinema Display to display the underlying map and a HTC G1 phone was used for the handheld AR interface as well as for the peephole interface. To estimate the pose of the phone in relation to the map the video from the phone is transmitted to a server via WiFi. The server analyses the video and estimates the phone's pose which is transmitted back to the phone. Image processing is performed at a rate of more than 8Hz. To select a POI the user has to tap it with the finger on the phone's screen. Once selected, the POI fades to grey.

5.2.2 Method of the Study

The experiment investigates the effect of visualizing off-screen objects using a handheld AR and a peephole interface. To make the experiment comparable with previous studies we use a design similar to [RSR$^+$07]. The participants had to select POIs on a map. The independent variables are the interaction technique (handheld AR or peephole interface) and the visualization technique (off-screen visualization or no off-screen visualization). The dependent variables are the tasks completion time, error rate, and the perceived task load measured using the NASA TLX [HS88].

The study consisted of two tasks: In the find-and-select task participants were asked to select the "greenest" POI displayed by the augmentation on the phone. The task consisted of six sub-tasks. The participant had to select the greenest POI out of 2, 4, 6, 8, 10, and 12 POIs. The POIs' colours were selected from a colour space spanning from green to red. The order of the sub-tasks was randomized. In the following select-in-order task participants were asked to select all POIs from the greenest to the reddest. There were sets with 2, 4, and 6 POIs resulting in three sub-tasks. The POIs' colours were also selected from a colour space spanning from green to red.

Both tasks were set up as a 2x2 repeated measurement design with interaction technique (Handheld AR or peephole interface) and off-screen visualization (with or without visualization of off-screen POIs) as the two independent variables. Participant performed both tasks with every condition. The order of the conditions was counterbalanced to reduce sequence effect. 12 persons (4 female) participated in the study. Most participants had a technical background and were aged between 22 and 38 years (mean age 30.42). For each sub-task the task-completion time was measured. In addition, it was recorded if a wrong POI was selected. To assess the participants' perceived task load we used questionnaires with the unweighted version of the NASA TLX.

Participants were first familiarized with the procedure, the setup of the study, and the NASA TLX questionnaire. Before each task a textual description of the task was given on the mobile phone's screen. By tapping the screen the participant started to perform the respective sub-tasks. Participants were not asked to provide feedback while performing the tasks. After finishing one task with one condition the participant filled the questionnaire. After finishing all tasks participants were debriefed.

Our hypothesis was that the off-screen visualization reduces the task completion time for handheld AR and the peephole interface. This hypothesis is based on the assumption that the off-screen visualization provides additional cues but does not interfere with the augmentation. However, we assumed that interpreting the visualization of off-screen

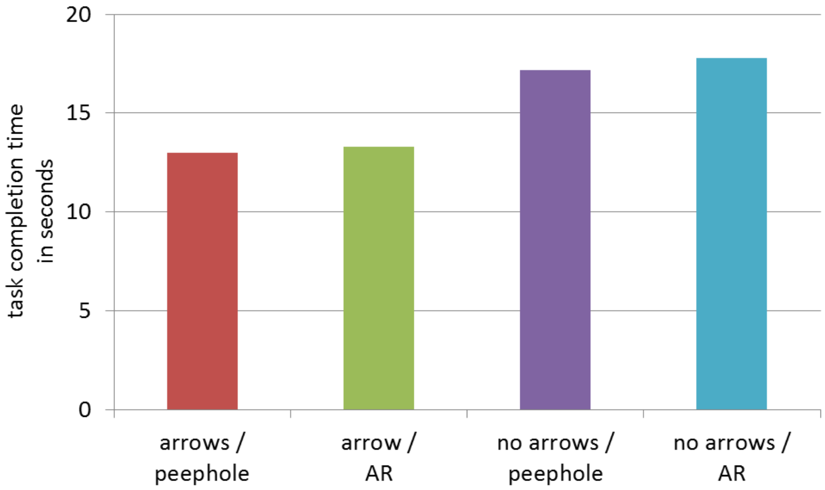


Figure 5.10: Average task completion time for the four conditions to test off-screen visualizations for printed maps in the find-and-select tasks.

objects increases the perceived task load in particular the mental demand. Regarding the differences between handheld AR and the peephole interface we expected only small effects that are consistent with the results reported by Rohs et al. [RSR$^+$07].

5.2.3 Results and Discussion

In the following we report the experiment's results. In total, participants made 19 errors in the first task and 13 errors in the second task but we did not find significant differences.

5.2.3.1 Completion Time

Figure 5.2.3.1 shows the average completion times for the find-and-select tasks. The time to select a target for the peephole interface with off-screen visualization was significantly faster than for handheld AR without off-screen visualization (p<.01, r=.47) and the peephole interface without off-screen visualization (p<.01, r=.46). Completion time for handheld AR with off-screen visualization was also significantly faster than for handheld AR without off-screen visualization (p<.01, r=.51) and the peephole interface without off-screen visualization (p<.01, r=.41).

Figure 5.2.3.1 shows the average time to select an item for the select-in-order tasks. Selection time for the peephole interface with off-screen visualization was significantly shorter than for handheld AR without off-screen visualization (p<.05, r=.10) and the peephole interface without off-screen visualization (p<.01, r=.22). Selection time for handheld AR with off-screen visualization was also significantly shorter than for handheld AR without off-screen visualization (p<.01, r=.12) and the peephole interface without off-screen visualization (p<.001, r=.23).

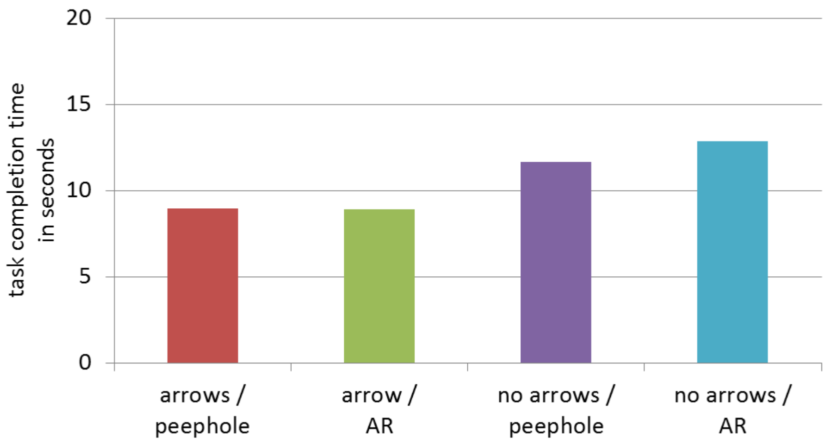


Figure 5.11: Average task completion time for the four conditions to test off-screen visualizations for printed maps in the select-in-order tasks.

5.2.3.2 Subjective Results

Analysis of the questionnaires showed that conditions with additional off-screen visualization were rated less demanding for both tasks. For the first task, the unweighted NASA TLX score for the peephole interface with off-screen visualization (M=4.53) was significantly lower than for handheld AR without off-screen visualization (M=5.74, p<.01, r=.77) and the peephole interface without off-screen visualization (M=6.49, p<.01, r=.71). The unweighted NASA TLX score for handheld AR with off-screen visualization (M=5.12) was also significantly lower than for handheld AR without off-screen visualization (p<.05, r=.75) and the peephole interface without off-screen visualization (p<.05, r=.66). We found a number of significant differences for the individual NASA TLX values that are all consistent with the result for the unweighted TLX score.

For the second task, the unweighted NASA TLX score for the peephole interface with off-screen visualization (M=5.36) was significantly lower than for handheld AR without off-screen visualization (M=7.04, p<.01, r=.79) and the peephole interface without off-screen visualization (M=7.47, p<.01, r=.83). The unweighted NASA TLX score for handheld AR with off-screen visualization (M=4.82) was also significantly lower than for handheld AR without off-screen visualization (p<.01, r=.75) and the peephole interface without off-screen visualization (p<.001, r=.88). We found a number of significant differences for the individual NASA TLX values that are all consistent with the result for the unweighted score.

5.2.3.3 Discussion

The experiment supports our hypothesis that a visualization of off-screen objects reduces the task completion time. Visualization of off-screen objects leads to a task completion time that is about one quarter lower. Furthermore, we also observed the time difference between handheld AR and the peephole interface is much smaller than the difference between off-screen visualization and no off-screen visualization. Unlike [RSS$^+$09] we did not find significant differences between the peephole interface and handheld AR. Based on our data we assume that our third hypothesis, that differences between handheld AR and the peephole interface are negligible for the performed tasks and setup, is also valid.

Contrary to our second hypothesis the perceived task load is lower if off-screen objects are visualized. Even though participants had to interpret the additional visualization it seems that the arrows assisted the participants more than we expected. In addition, the off-screen visualization might have reduced visual context switches between the phone's display and the physical map. Differences between the peephole interface and handheld AR regarding mental demand were not consistent for the two tasks.

Overall we observed the same trend for the subjective results and for the objective results. If off-screen visualizations are used participants were more efficient and they perceived the interface as more efficient. We assume that for tasks as those used in this study it is questionable if handheld AR provides a major benefit over the peephole interface. We can, however, not predict the effect of off-screen visualization in more complex tasks. If larger surfaces are used and the density of off-screen objects is increased the arrows used in this study will start to overlap which will clearly reduce the effectiveness of this particular visualization at a given point.

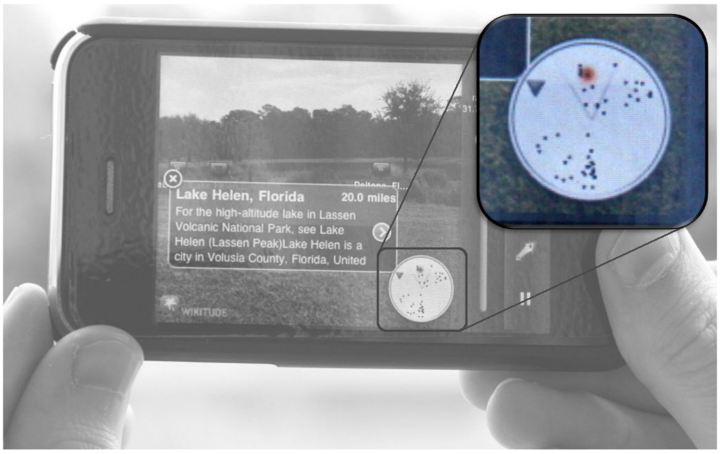
5.3 Off-screen Visualizations for POIs

As shown in the previous sections, providing an off-screen visualization provides a clear benefit for users when interacting with a 2D plane such as digital maps or if augmenting 2D physical maps. Handheld AR for interacting with physical object in a 3D environment provides an additional degree of freedom. A typical use case for handheld AR for 3D environments is the augmentation of POIs. We assume that determining the position of POIs that are connected with additional information is even more demanding compared to interacting with 2D content.

Current handheld AR applications (e.g. Wikitude or Layar) for tourists provide an overview about nearby sights using an additional mini-map, usually centred in the lower half of the display. Figure 5.12 shows the mini-map provided by Layar[4] a typical example of a commercial handheld AR application. The mini-map shows POIs in the users' surrounding using small dots. The user, or more precisely the phone, is located in the centre of the map. The viewing angle of the camera is indicated with two lines that originate in the centre of the map. If a POI is selected it is highlighted using a red shadow.

The augmented environment, the augmentation itself, and the mini-map, however, have different reference systems. Therefore, it can be assumed that interpreting the 2D mini-map and align it with the augmented environment demands high mental effort. We conjecture that an off-screen visualization directly embedded into the augmentation can reduce this mental effort. In this section, we therefore develop a 3D visualization of off-

[4] Website of the application Layar: `http://www.layar.com` last accessed 24 November 2011



Lake Helen, Florida 20.0 miles

For the high-altitude lake in Lassen Volcanic National Park, see Lake Helen (Lassen Peak)Lake Helen is a city in Volusia County, Florida, United

Figure 5.12: The off-screen visualization of the Android version of Layar. Layar displays nearby POIs using a mini-map.

screen objects for handheld AR applications. Three potential visualizations are designed in Section 5.3.1 and compared in Section 5.3.2. Based on the results the design is revised and evaluated using a mini-map as baseline in Section 5.3.3.

5.3.1 Developed Off-Screen Visualizations

Existing off-screen visualizations for mobile devices are designed for 2D applications. When augmenting POIs these objects are located in 3D and the visualization must be able to convey an additional degree of freedom. Therefore, the off-screen visualizations investigated in Section 5.1 and Section 5.2 cannot be directly applied. Therefore, we develop three arrow-based off-screen visualizations in the following. Figure 5.13 provides an overview about the three visualizations. Figure 5.13.a shows a scene from the top and Figure 5.13.b-d show the same scene from the user's perspective using the three visualization technique.

The three approaches are based on stretched arrows that communicate the distance between the user and the object by stretching the body of the arrow. The longer the distance the longer is also the body of the arrow. We decided for stretched arrows instead of scaled arrows. In a 3D environment determining the orientation of an arrow is more challenging than in 2D. The body of a stretched-arrow helps to determine the arrow's direction. In contrast, a scaled-arrow that has now body does not provide such a hint which makes identifying its direction more difficult.

2D Arrows

The first type of arrow is based on the assumption that 2D arrows are easier to perceive than 3D arrows. However, 2D arrows cannot point directly at 3D positions because

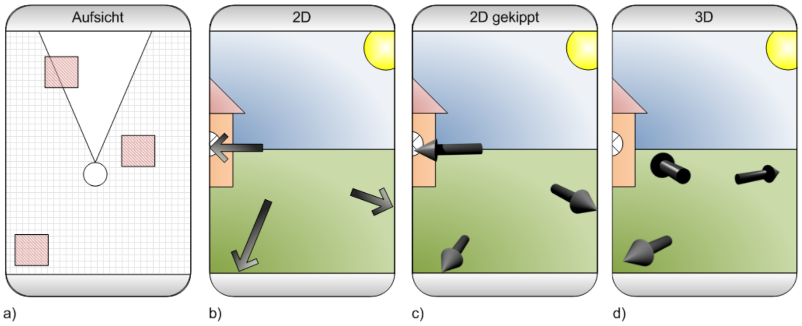


Figure 5.13: Concept of the three off-screen visualizations for POIs: a) Scene from top b) 2D Arrows, c) Tilted 2D Arrows, and d) 3D arrows.

they lack the additional dimension. In addition, not all possible 3D positions can be encoded. Therefore, this visualization is designed to only encode positions on the 2D plane around the user. The angle of a line between a position in front of the device and the object defines the position of the arrow's head at the left, bottom, or right border of the screen. If an object is located in the left-front or right-front just outside of the device's field of view the arrow is horizontally oriented. The arrow moves at the devices border if the object circles around the user.

Tilted 2D Arrows

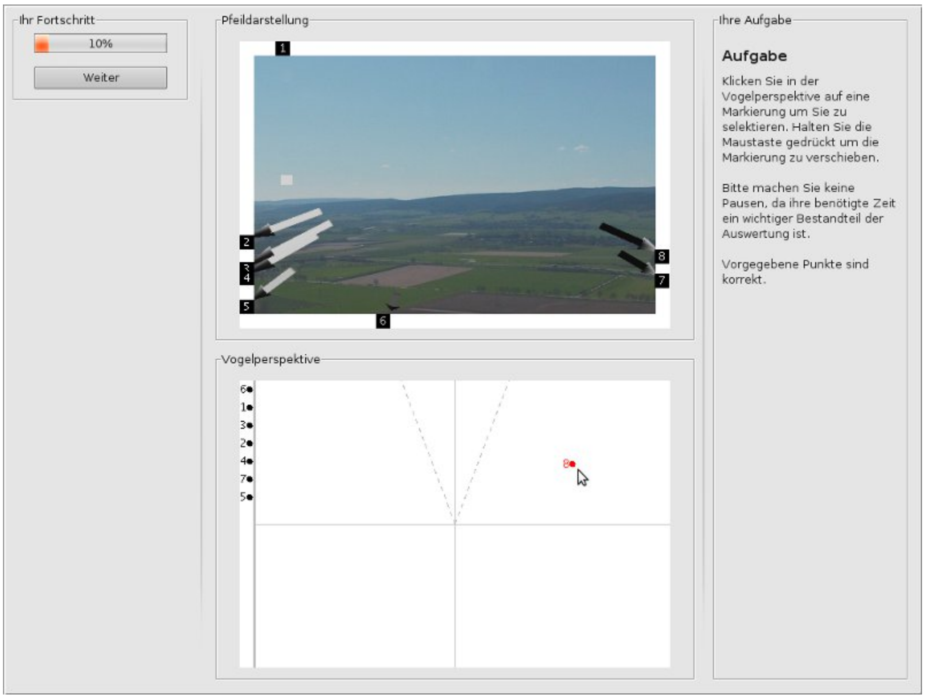
Pure 2D arrows could have the disadvantage that they tend to overlap. Especially if visualizing a larger number of object arrows in the bottom-left and bottom-right corner of the screen will intersect. To reduce this limitation the second off-screen visualization tilts the arrows inside the scene if the object is located behind the user. Apart from that this type of arrow behaves the same as the 2D arrows. To tilt the arrows in the scene they must, however, be rendered in 3D.

3D Arrows

To use the full potential of a 3D visualization the arrows must point directly at the objects. For the last off-screen visualization the arrows have the same orientation as the line between a position in front of the device and the object. To reduce overlapping of different arrows the arrows are positioned on a semicircle as long as the respective object is in front of the user. If the object is behind the user the respective arrow's head is located at the bottom of the screen.

5.3.2 Comparison of Design Alternatives

To compare the developed arrow-based off-screen visualizations we conducted a controlled experiment. The aims of the study were determining how precisely user can interpret the visualizations and to asses participants preferences. The study follows the



Aufgabe

Klicken Sie in der Vogelperspektive auf eine Markierung um Sie zu selektieren. Halten Sie die Maustaste gedrückt um die Markierung zu verschieben.

Bitte machen Sie keine Pausen, da ihre benötigte Zeit ein wichtiger Bestandteil der Auswertung ist.

Vorgegebene Punkte sind korrekt.

Figure 5.14: Screenshot of the evaluation tool used to test different off-screen visualizations.

approach used by Pielot et al. [PKB10] to compare different tactile encoding to present spatial information. To enable participants to specify locations independently without the bias of an investigator interpreting the participants' statements an online tool is used to conduct the experiment. The online tool presents a virtual scene containing off-screen objects to the participants (see Figure 5.14). Participants must move icons to the position shown by the respective off-screen visualizations on a 2D plane.

5.3.2.1 Developed Prototype

We developed a web application that guides participants through the study. Similar to a Wizzard, different views are presented to the participants. First of all, three websites explain the test environment. Participants proceed by clicking on the "weiter" (German for "next") button. Following the introduction an image of an AR scene is presented to the participants in the upper part of the dialog (see Figure 5.14). The image imitates an AR application on a smartphone. The scene contains a number of POIs. One of the off-screen visualization techniques is used to highlight the location of objects that are not visible. Participants' task is to move place holders of the POIs on the 2D plane presented in the lower part of the dialog. Participant should try to move the place holder to the position they expect the POI to be located. After positioning different configurations of POIs using one visualization technique, a set of questions is presented to the participants.

The application is controlled by a web server that delivers a website to the participants. The first three web pages introduce the study and the test environment. Afterwards, a web page with an embedded Java Applet presents the positioning task and the questionnaires. In order to use the website and the Applet an Internet connection and the Java Runtime Environment 1.5 (or higher) is required. Results are transmitted to our web server when the participant completed the tasks and filled the questionnaire for one visualization technique.

5.3.2.2 Method of the Study

In the experiment participants performed a single task to compare the three visualization techniques using a within-subject design. The only independent variable is the visualization technique with three levels (2D arrows, tilted 2D arrows, and 3D arrows). The order of the conditions was randomized to reduce sequence effects. The location of the POIs specified by the participants and participants' subjective ratings are used as dependent variables. The difference between a POI's position and the selected position is measured in three ways: Using the Euclidean distance between the selected position and the true position, the angular deviation between the two positions using the virtual location of the user as central point, and the difference of the distance between the user's virtual position and the two positions. Participants' subjective impression is assessed using five questions (e.g. "How easy was it to estimate directions?").

We invited participants to take part in the study by announcing it via social networks and the learning management system of the University of Oldenburg. The test environment was online for a duration of 2 weeks. In total 107 persons started to participate but only 22 persons completed all tasks. According to the 22 participant's self-reports they were between 17 and 57 years old. Their average age was 27.85 years (SD=8.20 years). 5 participants are female and 17 male.

5.3.2.3 Results and Discussions

Analysing the collected data we found that the off-screen visualization only had a significant effect on the objective measures but not on the subjective measures. An overview about the three objective measures is provided in Figure 5.15 and Table 5.1. An ANOVA shows that the visualization technique had a significant effect on participants' accuracy when estimating the POIs' positions (p<.01). As we used three conditions the significance levels for the follow-up t-tests are reduced to $0.5/3 = .01\overline{66}$ with a Bonferroni correction. Participants estimated the POIs' position more precisely using 3D Arrows compared to 2D Arrows or Tilted 2D Arrows (both p<.01). The difference between 2D Arrows and Tilted 2D Arrows is not significant (p=0.37). The independent variable had no significant effect on the accuracy of the estimated distances (p=0.13) or the angular deviation (p=0.16).

An overview about the five subjective measures is provided in Figure 5.16. Analysing the participants' subjective ratings using an ANOVA showed that the off-screen visual-

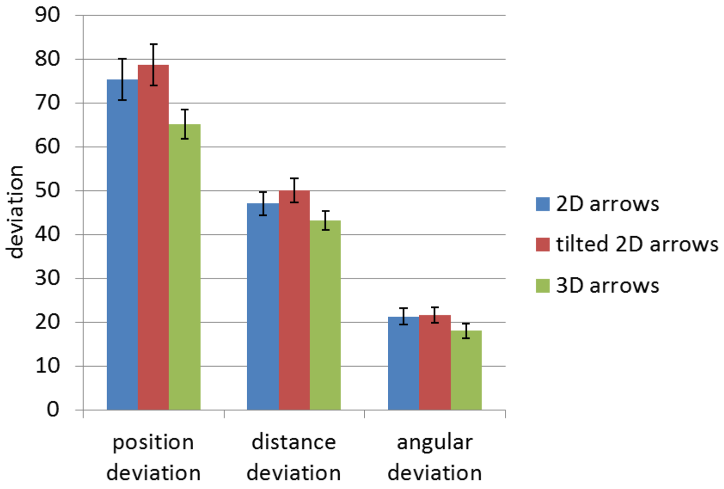


Figure 5.15: Deviation between the positions of visualized POIs and the positions selected by the participants (error bars show standard error).

Table 5.1: Average and standard deviation of the deviation between the positions of visualized POIs and the positions selected by the participants.

	2D Arrows	Tilted 2D Arrows	3D Arrows
position deviation	75.37 (SD=21.96)	78.66 (SD=21.97)	65.21 (SD=15.58)
distance deviation	47.01 (SD=12.62)	50.00 (SD=12.84)	43.14 (SD=9.96)
angular deviation	21.29 (SD=8.50)	21.62 (SD=8.25)	18.04 (SD=7.58)

ization had no significant effect. The individual p-values are: estimate direction p=0.54, estimate distances p=0.74, mental demand p=0.21, frustration p=0.98, success p=0.55. Nonetheless, it is remarkable that the mean of all subjective measures favour the 3D arrows.

The objective measures as well as the subjective measures show the same tendency. However, only participants' accuracy when estimating a POI's position is significantly higher when using 3D Arrows compared to the other visualizations. Therefore, we analysed the objective measures in more detail. We separated the data by considering POIs in the front and POIs in the back independently.

Figure 5.17 shows the objective measures when treating POIs in front of the user's position and POIs behind the user's position independently. As a series of t-tests is performed on the data a conservative level of significance is used to ensure the validity of the results. Thus, the means are only considered as statistically different if the probability of Type I errors is below 0.1% (p<.001). An ANOVA shows no significant effect on any of the measures for the POIs located behind the user's position (p>.22 for all measures). For estimating a POIs distance from the user's virtual position considering only POIs in front of the user, an ANOVA shows that the off-screen visualization also had no effect (p=.20).

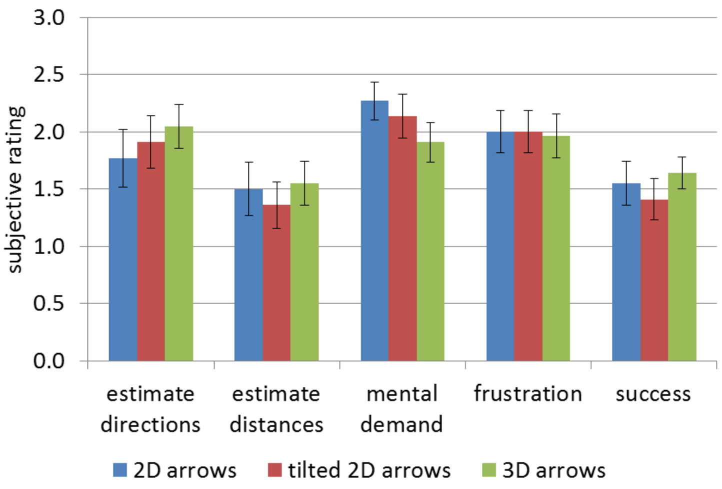


Figure 5.16: Subjective ratings of the three off-screen visualizations for POIs on a five point scale (0-4). Higher numbers mean better ratings (error bars show standard error).

For the POIs located in front of the user's position the visualization had a significant effect on the on participants' accuracy when estimating the POIs' angles (p<.0001). Pair wise t-tests show that the angular deviation is significantly lower using 3D arrows (M=13.69, SD=5.84) than using 2D arrows (M=22.21, SD=12.74, p<.001) or using tilted 2D arrows (M=21.49, SD=10.88, p<.0001). The difference between the 2D arrows and the tilted 2D arrows is not significant (p=.73). Similarly, an ANOVA shows that the visualization had a significant effect on participants' accuracy when estimating the POIs' position (p<.0001) for POIs located in front of the user's position. Pair wise t-tests show that the deviation is significantly lower using 3D arrows (M=57.72, SD=15.26) than using 2D arrows (M=84.25, SD=33.36, p<.0001) or using tilted 2D arrows (M=83.47, SD=32.85, p<.0001). Again, the difference between the 2D arrows and the tilted 2D arrows is not significant (p=.90).

The results show that the off-screen visualization technique had no significant effect on the participants' subjective rating. Even though, the average ratings suggest that there might be an advantage for the 3D arrows. The objective measures show that participants estimated the position of POIs more precisely using 3D arrows. The difference is, however, relatively small and we did not found a significant effect on the other objective measures. Separating the data by considering POIs located in front of the user's position and those located in the back shows that the visualization technique had only an effect if the POIs are located in the front. In this case, the 3D arrows significantly outperform the 2D arrows and the tilted 2D arrows.

The results suggest that the 3D arrows perform particularly well for positions located in front of the user's position. The experimental setup, however, does not allow a statistical comparison of the participants' performance for POIs located in front of the user's position and those located behind. Looking at the average deviations, however, it can be assumed that it might be easier to estimate positions in the front using 3D arrows than

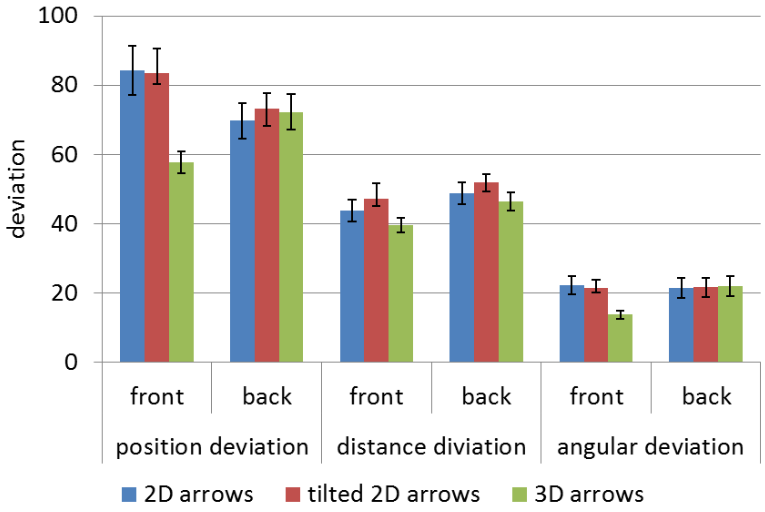


Figure 5.17: Deviation between the positions of visualized POIs and the positions selected by the participants after separating the POIs in front and in the back of the user's virtual position (error bars show standard error).

estimating locations in the back using 3D arrows. The 3D arrows treat positions in the front and positions in the back differently. If presenting frontal positions the arrows are located on a semicircle. If a POI is behind the user's position, however, the arrow's head is located at the bottom of the screen. An option to further improve the 3D arrows that seems obvious is to position the arrows that point at rear location also on a semicircle. Accordingly all arrows would be located on one ellipse. This might also improve the arrows' intuitiveness and reduce the required mental effort for interpreting them because all arrows are positioned consistently. Similar to previous work on 2D off-screen visualizations the findings are also limited because participants were not able to move the scene and therefore needs to be confirmed in a further study (see below).

5.3.3 Comparison of a revised design and a Mini-Map

In order to determine if an arrow-based off-screen visualization can compete with what is used by current commercial applications we conducted an according controlled experiment. In the experiment we compared the arrow-based visualization with a minimap. Mini-maps, using a variation of overview and detail, are well established means that have been investigated in context of different types of applications. They have been intensively studied for desktop applications [CKB08] and their performance has been demonstrated for browsing Web pages [RPKV06] and maps on mobile devices [RPKV06].

To compare mini-maps that are used by current commercial applications we conducted an according controlled experiment. Based on the results of the study described in the previous section the 3D arrows are further refined. In addition, we develop a sensor-

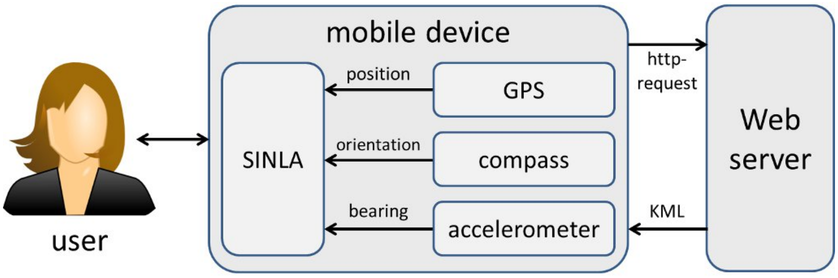


Figure 5.18: The architecture of the SINLA prototype that augments POIs.

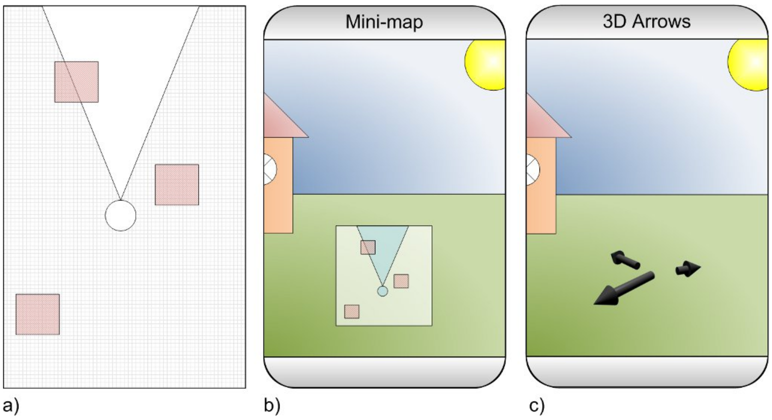


Figure 5.19: Concept of the two off-screen visualizations for POIs: a) Scene from top, b) mini-map, and c) 3D arrows.

based handheld AR application. The application augments the camera image by estimating the device's position using GPS, accelerometers and a compass as input. Using this application a field study with passersby on a public square is conducted.

5.3.3.1 Developed Prototype

We assume that a 2D map presented besides the augmented scene, as in Figure 5.19.b, demands effort to be interpreted. Based on the conducted study we refined the 3D arrows by arranging all arrows on a circle. The arrows point directly at nearby POIs. The centroid of each arrow is located on this circle and thus, all arrows are on the same plane. The centre of the circle is moved in front of the user's position to be inside the viewport. To reduce occlusion among the arrows the plane is slightly tilted towards the user. The arrows rotate according to the orientation of the phone, just like a compass with multiple needles. To present the distance between the viewer and the object the arrows are scaled in length according to this distance. The scale factor can be adjusted just like the zoom level in digital maps. As shown in Figure 5.19.c, arrows are not hidden if an off-screen object becomes an on-screen object (i.e. the object is visible inside the camera's video) to avoid confusing the user.

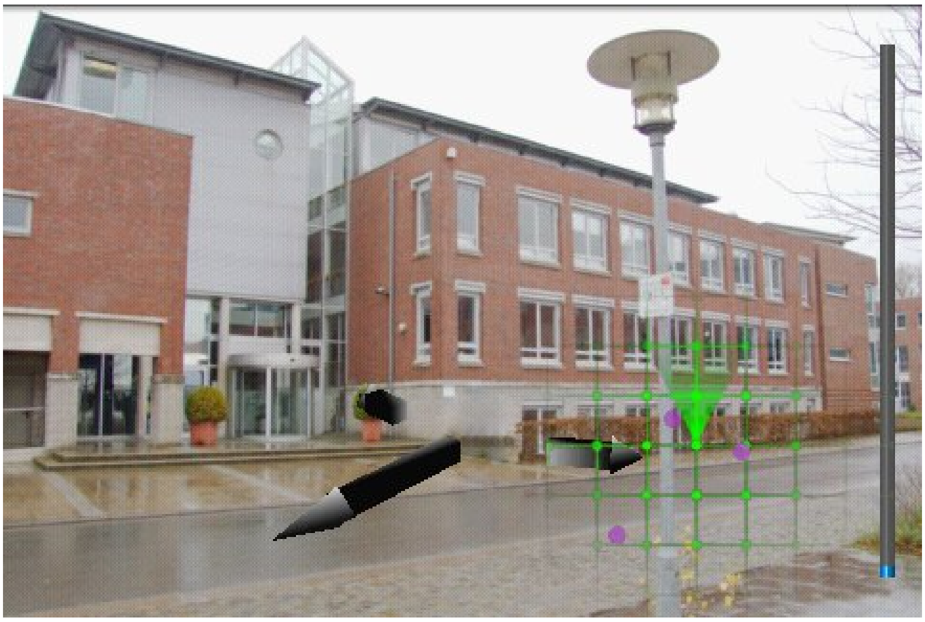


Figure 5.20: Screenshot with both off-screen visualization for POIs. Both visualization techniques are combined for illustration purpose. Only one was used at a time for the evaluation.

To compare the arrow-based visualization with the state-of-the-art we implemented a mini-map that also rotates with the orientation of the phone. A highlighted cone shows the area of the real world that is visible on the phone's display. To obtain comparable results the mini-map is centred at the same location with the same size as the arrows. To highlight POIs inside the camera's video we use circles that overlay the objects inside the physical scene. If a POI is near the centre of the display a short description is painted on top of a semi-ellipse connected to the circle by a thin line. The system was implemented for Android smartphones. A screenshot containing a side-by-side comparison of the application's two presentation techniques is shown in Figure 5.20.

5.3.3.2 Method of the Study

To compare the arrow-based visualization of nearby POIs with mini-maps we conducted a user study with the system described above. In the experiment participants performed two tasks using the system. The experiment's independent variable was the visualization technique used to show POIs located in the off-screen. In the control condition, participants used a mini-map and in the experimental condition they used the arrow-based visualization instead. The study consisted of two tasks. We used a repeated-measures design for both tasks. The tasked are always performed in the same order but the order of the conditions have been counterbalanced to reduce sequence effects.

For the first task the device displayed four virtual POIs randomly distributed around the user. Participants' task was to read the names of the POIs. In order to do that,

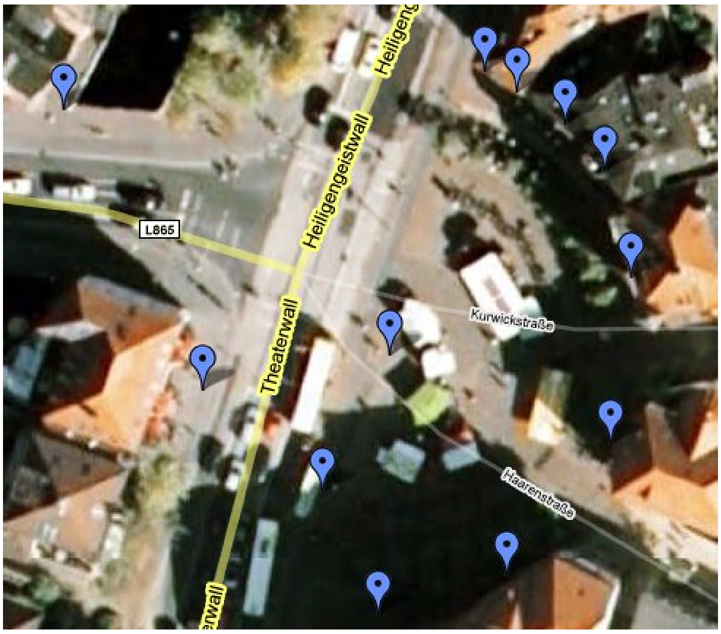


Figure 5.21: Top-down view on the square where the user study to compare an arrow-based off-screen visualization with a mini-map took place. The blue place marks show the position of the POIs used during the study.

they had to search for the POIs by turning around on the spot. A POI's name was written on the top of the screen if the POI was located at the centre of the display. The dependent variables were the time needed to read the names of all POIs and a rating of the visualization technique on a six point scale.

In the second task the device showed a set of four nearby POIs that are randomly selected from the 12 nearby POIs (e.g. buildings, shops, and a bus station) shown in Figure 5.21. Participants were asked to turn in a specific direction before starting the task and memorize the location of the POIs without turning around. After they finished memorizing, the device was removed, and participants had to tell which POIs were displayed. The dependent variables were the time needed to memorize the POIs and a rating of the visualization technique on a six point scale. In addition, the number of correctly estimated POIs and the difference between the named POIs and the displayed POIs in meters and angle were measured.

We set up the evaluation booth on the Julius-Mosen-Platz a public square in the city centre of Oldenburg – a medium size European city. Figure 5.22 shows a participants and one of the evaluators during the study. The study was conducted on a Saturday from 11.00 to 16.00. Two teams of experimenters guided participants through the tasks simultaneously. Passersby were randomly asked to participate in the study. After a person had agreed to participate in the evaluation, the experimenter made the participants familiar with the concept of presenting POIs using AR and the two visualization techniques. After conducting both tasks participants were interviewed to collect demographic informa-

Figure 5.22: Photo of an evaluator and a participant during the user study that compares an arrow-based off-screen visualization with a mini-map conducted at a public place.

tion. In addition we asked participants to self-assess their experience with smartphones, navigation skills, and experience with virtual reality (VE) on a six point scale.

We conducted the user study with 26 participants, 13 female and 13 male, aged 21-41 (M=22.4, SD=7.2). The subjects were passersby, so most of them were familiar with the local place. All subjects were volunteers, chosen without any selection by age, nationality or other criteria. None of them were familiar with the used application.

Our assumption was that participants are faster with the arrow-based technique because they do not have to mentally align two different reference systems. Thus, we also expected that participants understand the arrow-based technique faster and that this technique can be interpreted quicker. However, we assumed that participants can localize POIs more precisely with the mini-map because of their experience with map usage and because of the more abstract 2D visualization.

5.3.3.3 Results and Discussion

After conducting the experiment we collected and analysed the data. Participants completed the first task in 27.2s using the arrow-based technique and in 26.7s using the minmap. We did not find significant results (p=.47). Therefore, we refrain from a detailed discussion and focus on the results of the second task in the following. In addition to the differences between the visualization techniques, significant effects of gender and stated experience with virtual environments are also reported if applicable. Unless otherwise noted, a t-test is used to derive the p-values.

Using the arrow visualization subjects correctly identified significantly (p<.05) more POIs (M=2.2) than using the mini-map (M=1.6) (see Figure 5.23). In particular, males

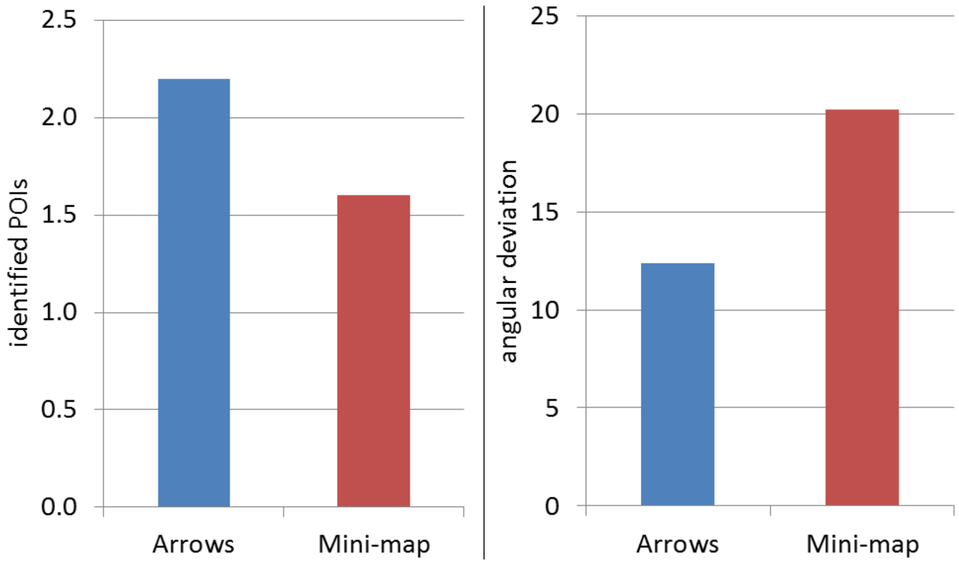


Figure 5.23: Average number of correctly identified POIs using a mini-map and the arrwor-based visualization (left) and the average angular deviation between the displayed POIs and the guessed POIs (right).

were significantly better (p<.05) when using the arrow-based method (M=2.1) in contrast to using the mini-map method (M=1.3) when identifying POIs. The difference for females was smaller. On average, they identified 2.3 POIs using the arrow-based technique and 1.9 POIs using the mini-map. Furthermore participants who stated to be good in VE were also significant better (p<0.05) when using the arrow method (M=2.4 compared to M=1.3).

To compare the positions of the displayed POIs with the positions stated by the participants the respective geo-coordinates were used. Using these geo-coordinates the deviation in meters between these positions were calculated. Using arrows the distance between the POIs' correct position and the guessed position was lower (M=18.0m) than using the mini-map (M=23.3m) but the difference is not significant. The difference of the distance from the user to the correct POI and the distance from the user to the guessed POI was smaller using arrows (M=29.9m) than using the mini-map (M=38.8m) but the difference was also not significant.

We calculated the angle between the displayed POIs and the guessed POI (see Figure 5.23). The angular deviation was significantly smaller (p<.05) using arrows (M=12.4°) than using the mini-map (M=20.2°). In particular, females profited from the arrows (M=9.7°) and were significant better with arrows (p<.05) than using the mini-map (M=18.2°). Also subjects who stated to be good in VE were significant better (p<.001) when using the arrows (M=8.5°) instead of the mini-map (M=25.2°).

On average participants are slightly faster using the arrows (M=21.4s versus M=24.1s) but no significant difference was found. The ratings are nearly equal with M=2.9 for the arrows and M=2.9 for the mini-map whereas 1.0 is best and 6.0 is worst.

We found some general tendencies that we, however, cannot prove to be significant. Using the arrows subjects that stated to have experience with virtual environments were on average 3.4s slower but identified 0.3 more objects correctly than subjects who stated to have little experience. In contrast, we found opposite results for the mini-map: Subjects that stated to have experience with virtual environments were on average 6.1s faster and identified 0.5 less objects correctly than subjects who stated to have little experience. Furthermore females were always better than males on average for both visualizations (e.g. 4.9° smaller angular deviation and 0.4 more objects correctly identified) but slower (3.3s) in all measured values.

We analysed our data to find anomalies in our sample of the population. Overall females were 3.8 years younger than males ($p<0.05$). Furthermore, on a five point scale (from 1=no skills to 5=highly skilled) females rated their navigation skills 0.7 points worse than their male counterparts ($p<0.0001$) and females rated their experience with virtual environments (again on a five point scale) 1.3 points worse than males ($p<0.0001$). On a five point scale (1=not familiar to 5=very familiar), younger participants rated their own competence worse in the categories familiarity with the environment ($p<0.0001$, $r=-0.391$), navigation skills ($p<0.0001$, $r=-0.417$), and experience with virtual environments ($p<0.021$, $r=-0.261$). An analysis of covariance (ANCOVA) showed that the younger females are not the cause why younger participants rated themselves worse in general because gender is not a covariate in this correlation.

No significant difference between the visualization techniques has been found for the first task. We assume that the data is affected by noise induced by participants' lack of training. Furthermore, the inaccuracy of the used phone's build-in compass certainly affected the results. The HTC G1 is one of the first phones that provides compass data using a magnetometer. The data from the magnetometer is noisy but, even more important, can be biased by 30° or more. We checked the accuracy before starting the experiment. After completing the experiment we checked the sensor reading again and observed a clear bias which must have affected the outcome. However, for the second task a systematic bias in the sensor reading has no effect on the measures and the arrows clearly outperform the mini-map. Participants were faster and more precise with the 3D arrows. In particular, the absolute amount of correctly identified objects was higher.

The results support our first assumption, that participants are faster with the arrow-based technique, because they do not have to align different reference systems. Due to the same reason we expected that users perceive the arrows as more intuitive, indicated by higher ratings. Surprisingly, all ratings were almost equal. One reason might be that subjects were naive users, which were more interested in handheld AR applications than in the compared visualization techniques. Our last assumption, that participants are more precisely when localizing POIs with the mini-map, was contradicted. We assume that these results are mainly caused by the improved visualization of directions the 3D arrows provide. Because of our experimental design we cannot estimate if the 3D arrows visualize distances more effectively.

We asked participants to self asses their competence with virtual environments because we expected that this competence has a direct effect on participants' skills to navigate and orientate in an AR. On average, females performed better but with a lower self-assessment of their experience with virtual environments. This implies that some males overrated their own competence.

We were surprised that, in particular, subjects who stated to have experience with virtual environments were supported by the 3D arrows. This is contradictory to the results from Burigat and Chittaro [BCG06, BC07], who showed that presenting a destination in virtual environments using a 3D arrow is especially suitable for inexperienced users. We identified two potential reasons causing this contradiction: Let participants rate their competence in virtual realities might be too imprecise, especially compared to the preselected experts Burigat and Chittaro used in their experiment. Furthermore it's unclear if the results of an experiment about stationary virtual environments are applicable to handheld AR.

5.4 Summary and Implications

The aim of this chapter was to investigate techniques that enable users to determine the availability and location of physical objects that can serve as anchor for digital information using handheld AR. Through five user studies we compare existing visualization techniques for digital maps, investigate the impact of an off-screen visualization for augmenting physical maps, and develop an off-screen visualization for POIs. In the following we summarize the conducted studies and the developed visualizations. Furthermore, we outline the implications of our findings on the design of prospective handheld AR applications.

5.4.1 Summary

To base our work in the domain of off-screen visualization for handheld AR on solid ground we started with an investigation of existing off-screen visualization techniques for digital maps on mobile devices. In two user studies, that we conducted by publishing apps in the Android Market, we compared three off-screen visualizations for digital maps on mobile devices. In total our prototypes got installed over 12,000 times and we analysed the app usage contributed by over 5,000 installations. We show that the arrow-based off-screen visualizations proposed by [BC07] outperform the technique Halo developed by [BR03] for interactive tasks that involve the visualization of 20 or more objects. In addition, the two studies are the first controlled experiments that have been conducted by publishing an application in a mobile application store. Following this initial work we [HPP$^+$11, PHB11, HRB11, HRB12] and others [MBMC11, PHZB12, PPHB12] further used this approach.

Using an arrow-based off-screen visualization technique as basis we investigate its effect on handheld AR for physical maps. In a conducted controlled experiment we compare handheld AR with dynamic peephole interaction and analyse the effect of an off-screen visualization on both interaction techniques. We provide evidence that the off-screen visualization reduces the task completion time and reduces the perceived task load. It is argued that for tasks as those used in the study it is questionable if handheld AR provides a major benefit over the peephole interface.

As we showed that an off-screen visualization provides a clear benefit for digital maps and augmented physical maps we developed an off-screen visualization for 3D environments. Three techniques that visualize off-screen POIs in handheld AR applications are developed. Comparing the developed visualization we found only small overall differences. A more detailed analysis showed, however, that the developed 3D arrows outperform the other techniques for POIs in front of the user's position. The design is accordingly refined by using the 3D arrows' behaviour for frontal positions also for POIs located behind the user. To test the resulting design we developed a handheld AR application that uses a mobile phone's GPS receiver, compass, and accelerometer to estimate the phone's position and orientation. Using this system the developed off-screen visualization is compared with a mini-map used by current commercial applications in a controlled experiment. The results support the assumption that users are faster with the arrow-based technique and can localizing POIs with more precision.

5.4.2 Implications

The results of the studies investigating off-screen visualizations for digital maps confirmed that arrow-based visualizations are superior in terms of users' speed. In contrast to previous work, that only investigated static maps in highly controlled environments, our results show that arrows are also superior 'in the wild', with natural interaction contexts, and interactive maps. We can assert that this visualization has the potential to improve the performance of users interacting with common mobile map applications such as Google maps. Our results thus imply that we can improve the usability of one of the most common tasks supported by smartphones using an arrow-based off-screen visualization.

We also showed that an off-screen visualization can clearly improve the usability of handheld AR systems for physical maps. Our results enable to contrast this difference with the difference between handheld AR and dynamic peephole interaction. In line with previous work, our results indicate that handheld AR outperforms a dynamic peephole interface. The much larger advantage of an additional off-screen visualization, however, enables questioning if the difference between handheld AR and dynamic peephole interface is relevant. Our results imply that developers and designer should question if using handheld AR instead of dynamic peephole interfaces is really worth the hassle. At least, for similar handheld AR systems that enable the interaction with large surfaces, an off-screen visualization is likely the most crucial aspect to consider.

Using an iterative process we developed an off-screen visualization for POIs and similar objects that are distributed in 3D. The developed visualizations can serve as a blueprint for providing off-screen visualizations for visualizing geographic entities. We showed that the developed visualization outperforms mini-maps, the technique of choice in current commercial systems. Thus, we can conclude that using our development has the potential to clearly improve the usability of the currently most widely used handheld AR use cases. Applications, such as Layar, Wikitude, and Goolge Goggles, could easily integrate our work and improve the performance of millions of users.

Our work also has implications beyond the domain of handheld AR. To our knowledge, we conducted the first controlled experiments using mobile application stores as a research tool [HB10c, HPB10]. We demonstrated a new way to determine scientific evidence for mobile HCI research [HPP$^+$11]. Our work about off-screen visualizations pathed the way for a number of studies that involved more than hundred thousand participants [PHB11, HB11a, HRB11]. Together with work that showed how to collect qualitative feedback from mobile users in the large [MMB$^+$10], we opened a new door for conducting mobile HCI research.

6 Conclusions

In the conclusion of this thesis, we provide a summary of the research presented in this work in Section 6.1. Section 6.2 outlines the contribution of the presented research and we summarize guidelines that derived from our iterative development and evaluation in Section 6.3. We close this thesis with highlighting open issues and directions for future work in Section 6.4.

6.1 Summary

We introduce our work and present open research challenges for camera-based mobile interaction with physical objects in Chapter 1. Chapter 2 provides a classification of research in the field of 'Mobile Interaction with the Real World'. Along this classification we present related work in the areas of touch-based interaction, Point & Shoot, Continuous Pointing, and handheld AR for mobile interaction with physical objects. The chapter is concluded by a discussion and a summary of open research questions.

In Chapter 3, three camera-based interaction techniques to access information connected to physical objects are analysed. We describe the interaction techniques Point & Shoot, Continuous Pointing, and handheld AR as well as manually typing a URL using a virtual keyboard, which serves as the control condition. In a controlled experiment we compare Point & Shoot to access information about physical posters and manually typing a URL using a virtual keyboard. In two participatory user studies it is investigated how Point & Shoot is used to interact with printed photo photobooks and how Continuous Pointing is used to interact with posters. Based on the gained insights, interfaces that use the three camera-based interaction techniques are designed and implemented. The resulting interfaces are compared in a controlled experiment and it is shown that handheld AR outperforms the other techniques in terms of user preferences and perceived task load.

Chapter 4 investigates the design of handheld AR user interfaces for physical media. Two concrete use cases are selected in order to develop potential user interfaces. Printed photo books and music CDs are used as exemplary types of physical media. The same explorative participatory design process is used for both use cases. In two user studies, we collect information and functions that should be supported and ask participants to sketch interface designs for the respective handheld AR application. The proposed designs are consolidated and implemented as software prototypes. To compare the developed designs a controlled experiment is conducted for each type of media. The experiments show the effect of different alignments on acquiring information and on interacting with the augmentation.

Approaches that enable users to determine the availability and location of physical objects that can serve as anchor for digital information using handheld AR are investigated in Chapter 5. Three different off-screen visualization techniques for digital maps are compared in two controlled experiments. Based on the findings we adapted an off-screen visualization for handheld AR interaction with paper maps and evaluate its effect in a controlled experiment. Similarly, we design and implement an off-screen visualization for augmenting POIs using sensor-based handheld AR. To determine the effect on the users' speed, accuracy, and ability to identify POIs the visualization is compared with a mini-map in a user study.

6.2 Contribution and Results

The design of the user interface is crucial for usable systems and a high user experience. Although considering the users' functional and non-functional requirements restricts the design space, the interface design still offers a large number of design decisions. This, of course, also applies to applications that use camera-based interaction techniques. This thesis main contribution is the investigation of camera-based mobile interaction with physical objects from different perspectives. The contributions, also discussed in detail in Section 1.2, can be summarized as follows:

Comparison of Techniques for Mobile Interaction with Physical Objects

- Defining the research field described by the notion 'Mobile Interaction with the Real World', providing a categorization, and an in-depth discussion of previous work.
- Providing evidence that Point & Shoot is improves users' objective and subjective performance compared to typing an URL using a virtual keyboard [PHN$^+$08].
- Designing, implementing, and evaluating the three camera-based interaction techniques Point & Shoot, Continuous Pointing, and handheld AR. Show that handheld AR provides a significantly higher interaction satisfaction than the other interaction techniques [HB12].
- Improving existing algorithm for server-based Point & Shoot [HB08] as well as Continuous Pointing and handheld AR on mobile phones [HSB09] by increasing the speed and the number of objects than can be recognized.

Design Principles for On-Screen Content and Controls in Handheld Augmented Reality

- Designing, implementing, and evaluating handheld AR interfaces to provide a baseline for other researcher. In addition the findings and resulting interfaces can serve as a guideline and reference for interface designers [HB10a, HB11c].
- Providing evidence that a handheld AR interface should display information connected to a physical object in the reference system of the object [HB11c].
- Providing evidence that input controls for controlling services connected to physical objects in handheld AR applications should be aligned with the screen and remain in the reference system of the phone [HB11c].

Interface Designs for Off-Screen Visualization in Handheld Augmented Reality

- Conducting the first controlled experiment by publishing an application to a mobile application store. Thereby, previously proposed off-screen visualizations are revised and scientific evidence is provided that an arrows-based off-screen visualization outperforms other techniques for interactive tasks [HB10c, HPB10, HPP^{+}11].

- Showing that highlighting objects beyond the display with an off-screen visualization for handheld AR significantly improves the users' subjective and objective performance [HB10b].

- Designing and implementing an arrow-based off-screen visualization that highlight objects beyond the display for interaction with large objects using handheld AR. Found evidence that the arrow-based off-screen visualization significantly improves users' objective performance compared to providing an overview using mini-maps [SHB10].

6.3 Guidelines

Through the development of 13 prototypes that use camera-based mobile interaction techniques and 13 user studies we can derive a number of guidelines that can help developers to design such applications. In the following we highlight the most important guidelines that developers should consider and provide references to the respective section of this thesis that provides further information and validation.

Choose an Appropriate Interaction Technique

Choosing an interaction technique is the most fundamental decision. We identified the four interaction techniques Touching & Hovering, Point & Shoot, Continuous Pointing, and handheld AR to interact with physical objects (Sec. 2.1). Touching & Hovering is only appropriate if the user can get in direct contact to the object. The camera-based techniques enable interaction even over a distance. Developers should avoid using Continuous Pointing because the provided feedback is counterproductive (Sec. 3.5). One should choose handheld AR because it is preferred by users and reduces the perceived task load (Sec. 3.5). Point & Shoot should be considered as an alternative if handheld AR is too computationally demanding for a particular task.

Consider the User's Behaviour when Developing the Underlying Algorithms

Developing the required algorithms for camera-based interaction needs to be informed by users' real behaviour. Using Point & Shoot, for example, users take images that are often blurry and cover only a fraction of the photographed object (Sec. 3.3). Still, the way users take images can be nearly optimal for particular algorithms (Sec. 3.5). Using handheld AR it needs to be considered that users move the handheld device rapidly and that tracking objects is required if recognizing objects is too slow (Sec. 3.5.1).

Reduce the Search Space

A number of use cases require recognizing a very large number of physical objects, an amount that is far too large to be used in handheld AR applications today. To implement such use cases it is required to reduce the number of potential objects by multiple magnitudes. As we proposed for printed photo books, a viable approach is to reduce the number of potential objects by asking the user to select a natural subset beforehand (Sec. 3.3). Other use cases can require other approaches to reduce the search space. An application for interacting with advertisement posters should, for example, only consider posters that are currently used in the user's surrounding.

Align the Augmentation to the Objects for Fast Browsing

Handheld AR is can be considered as a browser that uses physical objects as anchors. Applications should therefore enable simple and fast browsing of available content. Aligning the augmentation to the augmented object enables to quickly get an overview about available content and to naturally focussing on particular objects (Sec. 4.3 and Sec. 4.4). Information about an object should be attached to this objects and information about a particular part of an object should be attached to this part. A large number of objects might, however, require clustering the objects according to their properties.

Avoid Touching in the Reference System of Augmented Objects

As the information presented with handheld AR should be aligned to the object it seems like a natural choice to also align input controls to the object. Align input controls to the object requires the user to touch in the reference system of the augmentation. As this is a difficult and error prone task it should be avoided (Sec. 4.3 and Sec. 4.4). Otherwise, users would have to focus on the physical object, its augmentation, and to touch it – all at the same time. One approach to avoid forcing the user to touch in the reference system of the augmentation is to present input controls at a fixed position on the phone's border for the object that is in the centre of the display (Sec. 4.2.3). This approach enables to very quickly select an object and still enables to easily use the controls. Another approach is to present controls on a separate view that is disconnected from the augmentation (Sec. 4.2.5). This should only be preferred if the required input is complex or lengthy.

Minimize the Screen Space Covered by the Augmentation

Handheld AR applications augment the reality using a handheld device. Not surprisingly, it is the combination of augmentation and reality. It is necessary to keep the augmentation minimal to enable users to easily identify and understand the physical context of the augmentation (Sec. 4.2). Therefore, applications need to avoid hiding objects behind the augmentation and keep the amount of occupied screen space minimal.

Communicate the Status of the System

Understandability is a fundamental requirement for interactive systems. This includes communicating the system's status to the user and it even extends to communicate the status of the computer vision algorithms in handheld AR applications. Continuous Pointing fails because it hides the algorithm's status and abilities from the user (Sec. 3.5). Handheld AR systems should provide the required feedback by showing the user the objects that are currently recognized and tracked. Highlighting recognized objects is a very simple approach to not only indicate which objects are interactive but also naturally communicates the system's status. Successful approaches to highlight objects are simply drawing a border around them and to grey-out the background (Sec. 4.2).

Make Objects Easily Discoverable

To interact with an object, it is required to discover the object first. Thus, it is essential to make physical objects and interactive parts of physical objects discoverable. Users that search for interactive regions of an object will likely scan it using a systematic pattern (Sec. 3.4) and the basic gestalt principles should therefore be applied when designing physical objects that contain multiple interactive regions. Outliers, interactive regions that are disconnected from others, should therefore be avoided.

Provide Means to Discover Objects that are not in the User's Focus

If the developer cannot define the distribution of the physical objects, as we cannot define how POIs are arranged on a map, additional means are required to communicate the position of objects that are not in the in the focus of the handheld device. Highlighting the position of physical objects can significantly increase the user's performance and the usability of the system (Sec. 5.2). Off-screen visualisations can communicate the relative position of interactive regions. 2D arrows that communicate objects relative direction and distance should be used to point at objects distributed on a flat surface (Sec. 5.2) and 3D arrows should be used for objects distributed around the user (Sec. 5.3).

6.4 Future Work

A number of open research questions and unsolved challenges were identified while conducting the research presented in this thesis. In the following we highlight the most promising directions for future work and suggest potential approaches.

6.4.1 Long-Term Usage and Fatigue

Handheld AR, as well as other camera-based interaction techniques, requires holding the phone in a certain posture. This posture is very similar to holding a camera phone to take a photo or taking a photo with a digital camera. Therefore, it might be assumed that holding a phone in this way is sufficiently comfortable for the user, at least for a short interaction time. In addition, during none of the studies described in this thesis we observed any indication of fatigue and to our knowledge no related work describes such observation. The duration of the studies described in this thesis are, however, only between 15 minutes and 60 minutes and the interaction has been interrupted by short breaks. Furthermore, related work also describes only studies with a similar time span.

The posture required for handheld AR is very constrained. No long-term study has been conducted that analyses users' long-term performance and preference. Future work, thus, needs to study the effect of long-term usage. We would expect that users are not willing to continuously use a camera-based interaction technique for a long period of time. Since interaction with mobile applications is fragmented [OTRK05] and mobile applications are often used for a short duration [BHSB11] it is unclear if this has to be considered a general limitation. In any case, future work should investigate how long-term fatigue can be considered in the design of the user interface. A promising approach for long-term studies in a natural usage context is publishing research prototypes in mobile application stores. As we showed in [HB10c, HPB10, HPP$^+$11, HRB11, HRB12] this approach can be used to collect very large amounts of objective data and as shown by [MMB$^+$10] can also be used to collect subjective feedback.

6.4.2 Social Acceptability

This work focus on the interaction between a single user, a handheld device, and multiple physical objects. Today, mobile phones are the most powerful and common handheld device and since mobile phones are personal devices it can be expected that each user has his/her own device. Thus, designing for single user interaction is the natural starting point when investigating the interaction design for mobile camera-based systems. Usage of interactive systems does, however, not happen in a vacuum. Cultural background, persons in the surrounding, and the peer group influences not only if one wants to interact but also how a person interact with a system [RB10]. Few related work actually exists that helps to understand the social acceptability of the interaction techniques addressed in this thesis.

We cannot be certain about the populations' long-term adoption of specific technologies and, as it can currently be observed for mobile phones, perception of technologies change over time and differs across cultures [RK03, Cam07]. Still, field work in related areas [GAR$^+$09, RB10, CWG$^+$11] suggest that an understanding of current social acceptability is important to shape research and development. In particular, pointing a phone at an object reveals a user's interest in this specific object. Users might be unwilling to explicitly reveal their interest in particular objects. Field work is necessary that investigates the effect of revealing one interest to specific objects in the surrounding.

6.4.3 Scalable and Robust Algorithms

The algorithms that are available today are adequate to study camera-based interaction techniques. A large corpus of work (see [WRM$^+$08, WSB09]), including our own contribution [HB08, HSB09], advanced the state of the art towards handheld AR using current smartphones. Still, fragile network connections, limited processing power, and low memory makes handheld AR especially challenging on mobile devices. Today, markerless content-based handheld AR on mobile devices is therefore limited to a few hundred objects. This already enables a number of use cases, in particular if requiring the user to select a subset of objects to narrow down the number of potential objects as we proposed in [HB08]. Still, developing handheld AR applications that enable to augment all printed books available, all commercial music CDs, or all buildings and monuments in the world, is not feasible today.

A number of groups work on algorithms to increases the amount of objects that can be recognized using handheld AR. A promising direction is incorporating the users' context in the recognition process. Just as we ask the user to select a subset of objects to narrow down the search space the usage context can also be used to automatically narrow down the number of potential objects. An example for this approach is using the users' location to only consider features in the users' proximity [TCG$^+$08]. Future work should further investigate other factors, such as time of the day, users' interests, and smartphones' additional sensors to improve the scalability and robustness of algorithms for handheld AR.

6.4.4 Considering Technical Limitations in the Design

The interfaces that we designed and implemented in this thesis can serve as role models for future applications. Their suitability has been demonstrated through a number of user studies and different use cases. Still, concrete technical solutions and algorithms for camera-based interaction with physical objects have specific characteristics that restrict the possible interaction techniques and interface designs. The physical interaction radius of handheld AR systems, for example, is limited by the recognition algorithm's ability to recognize small and large objects in a camera image. Latency when recognizing objects and the tracking rate further affects the interaction.

New algorithms and refined implementations will lead to improved technical solutions in the future. It will, however, not be possible to completely eliminate all limitations. Reducing, for example, the latency below humans' perception threshold for certain real-life systems might not be feasible. To reduce the effect on the user experience it might be beneficial to inform users about the system's state. A system could, for example, inform the user that a potential object has been detected even before it can be recognized. This could reduce the perceived latency without requiring improved algorithms. Future work should therefore investigate how technical limitations can be diminished by explicitly incorporating them in the interface design.

List of Figures

List of Tables

Bibliography

[ABPW07] Y. Anokwa, G. Borriello, T. Pering, and R. Want. A user interaction model for NFC enabled applications. In *Proceedings of the Workshop on Pervasive RFID/NFC Technology and Applications*, pages 357–361, 2007.

[ADS10] M. Alessandro, A. Dünser, and D. Schmalstieg. Zooming interfaces for augmented reality browsers. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 161–170, 2010.

[AH12] D. Ahlers and N. Henze. ¿donde esta?–surveying local search in honduras. In *Proceedings of the Workshop on Mobility and Web Behavior*, 2012.

[Arn06] T. Arnall. A graphic language for touch-based interactions. *Proceedings of the Workshop on Mobile Interaction with the Real World*, pages 18–22, 2006.

[Atk08] C.B. Atkins. Blocked recursive image composition. In *Proceeding of the International Conference on Multimedia*, pages 821–824, 2008.

[BAB$^+$07] S. Boring, M. Altendorfer, G. Broll, O. Hilliges, and A. Butz. Shoot & copy: phonecam-based information transfer from public displays onto mobile phones. In *Proceedings of the Joint International Conference on Mobile Technology, Applications, and Systems and the International Symposium on Computer-Human Interaction in Mobile Technology*, pages 24–31, 2007.

[BBCTSG05] A. Bernheim Brush, T. Combs Turner, M. A. Smith, and N. Gupta. Scanning objects in the wild: Assessing an object triggered information system. In *Proceeding of the International Conference Ubiquitous Computing*, pages 305–322, 2005.

[BBRS06] R. Ballagas, J. Borchers, M. Rohs, and J.G. Sheridan. The smart phone: a ubiquitous input device. *IEEE Pervasive Computing*, 5(1):70–77, 2006.

[BC07] S. Burigat and L. Chittaro. Navigation in 3D virtual environments: Effects of user experience and location-pointing navigation aids. *International Journal of Human-Computer Studies*, 65(11):945–958, 2007.

[BCG06] S. Burigat, L. Chittaro, and S. Gabrielli. Visualizing locations of off-screen objects on mobile devices: a comparative evaluation of three approaches. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 239–246, 2006.

[BDLR+07] G. Broll, A. De Luca, E. Rukzio, C. Noda, and P. Wisner. Mobile interaction with the real world. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 295–296, 2007.

[Bea08] D. Beaver. 10 billion photos. `http://www.facebook.com/note.php?note_id=30695603919`, 2008. Facebook Engineering Blog. October 15th.

[BETVG08] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, 2008.

[BFO92] M. Bajura, H. Fuchs, and R. Ohbuchi. Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. *ACM SIGGRAPH Computer Graphics*, 26(2):203–210, 1992.

[BH10] G. Broll and D. Hausen. Mobile and physical user interfaces for NFC-based mobile interaction with multiple tags. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 133–142, 2010.

[BHP+08a] G. Broll, M. Haarländer, M. Paolucci, M. Wagner, E. Rukzio, and A. Schmidt. Collect&Drop: A technique for multi-tag interaction with real world objects and information. In *Proceedings of the Conference on Ambient Intelligence*, pages 175–191, 2008.

[BHP+08b] G. Broll, M. Haarländer, M. Paolucci, M. Wagner, E. Rukzio, and A. Schmidt. Collect&Drop: A Technique for Physical Mobile Interaction. In *Proceedings of the International Conference on Pervasive Computing*, pages 103–106, 2008.

[BHSB11] M. Böhmer, B. Hecht, J. Schöning, and G. Bauer. Falling asleep with angry birds, facebook and kindle–a large scale study on mobile application usage. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, 2011.

[BKHB09] G. Broll, S. Keck, P. Holleis, and A. Butz. Improving the accessibility of NFC/RFID-based mobile interaction through learnability and guidance. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, 2009.

[BKM09] M. Billinghurst, H. Kato, and S. Myojin. Advanced interaction techniques for augmented reality applications. *Virtual and Mixed Reality*, pages 13–22, 2009.

[BLM02] M. Beaudouin-Lafon and W. Mackay. Prototyping tools and techniques. *The Human-Computer Interaction Handbook*, pages 1006–1031, 2002.

[Blu09] D. Blum. Bild- und Gestenerkennung zum Steuern eines MP3-Players mit einem N95. Bachelor thesis, Media Informatics and Multimedia Systems Group, University of Oldenburg, 2009.

[Blu10] D. Blum. Handheld Augmented Reality zur Interaktion mit gedruckten Fotos. Diploma thesis, Media Informatics and Multimedia Systems Group, University of Oldenburg, 2010.

[BR03] P. Baudisch and R. Rosenholtz. Halo: a technique for visualizing offscreen objects. In *Proceedings of the International Conference on Human Factors in Computing Systems*, pages 481–488, 2003.

[BRSB05] R. Ballagas, M. Rohs, J.G. Sheridan, and J. Borchers. Sweep and Point & Shoot: Phonecam-based interactions for large public displays. In *Proceedings of the International Conference on Human Factors in Computing Systems*, 2005.

[BSR+07] G. Broll, S. Siorpaes, E. Rukzio, M. Paolucci, J. Hamard, M. Wagner, and A. Schmidt. Comparing techniques for mobile interaction with objects from the real world. In *Proceedings of the Workshop on Pervasive Mobile Interaction Devices*, 2007.

[BTVG06] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. In *Proceedings of the European Conference on Computer Vision*, pages 404–417, 2006.

[Cam07] S.W. Campbell. A cross-cultural comparison of perceptions and uses of mobile telephony. *New Media & Society*, 9(2):343, 2007.

[CB06] X. Cao and R. Balakrishnan. Interacting with dynamically defined information spaces using a handheld projector and a pen. In *Proceedings of the symposium on User interface software and technology*, pages 225–234, 2006.

[CDF+05] K. Cheverst, A. Dix, D. Fitton, C. Kray, M. Rouncefield, G. Saslis-Lagoudakis, and J.G. Sheridan. Exploring mobile phone interaction with situated displays. In *Proceedings of the Workshop on Pervasive Mobile Interaction Devices*, 2005.

[CDN88] J.P. Chin, V.A. Diehl, and K.L. Norman. Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of the International Conference on Human Factors in Computing Systems*, pages 213–218, 1988.

[CKB08] A. Cockburn, A. Karlson, and B.B. Bederson. A review of overview+ detail, zooming, and focus+ context interfaces. *ACM Computing Surveys*, 41(1):2, 2008.

[CKH07] C. Cheong, D.C. Kim, and T.D. Han. Usability evaluation of designed image code interface for mobile computing environment. *Human-Computer Interaction. Interaction Platforms and Techniques*, pages 241–251, 2007.

[CLD$^+$10] S. Carter, C. Liao, L. Denoue, G. Golovchinsky, and Q. Liu. Linking Digital Media to Physical Documents: Comparing Content-and Marker-Based Tags. *IEEE Pervasive Computing*, 9(2):46–55, 2010.

[CM92] T. P. Caudell and D. W. Mizell. Augmented reality: An application of heads-up display technology to manual manufacturing processes. In *Proceedings of the International Conference on System Sciences*, pages 659–669, 1992.

[CPV09] I. Cappiello, S. Puglia, and A. Vitaletti. Design and Initial Evaluation of a Ubiquitous Touch-Based Remote Grocery Shopping Process. In *Proceedings of the International Workshop on Near Field Communication*, pages 9–14, 2009.

[CS09] K. Church and B. Smyth. Understanding the intent behind mobile information needs. In *Proceedings of the International Conference on Intelligent user interfaces*, pages 247–256, 2009.

[CWG$^+$11] L.G. Cowan, N. Weibel, W.G. Griswold, L.R. Pina, and J.D. Hollan. Projector phone use: practices and social implications. *Personal and Ubiquitous Computing*, pages 1–11, 2011.

[DCDH05a] N. Davies, K. Cheverst, A. Dix, and A. Hesse. Understanding the role of image recognition in mobile tour guides. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 191–198, 2005.

[DCDH05b] N. Davies, K. Cheverst, A. Dix, and A. Hesse. Understanding the role of image recognition in mobile tour guides. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 191–198, 2005.

[DCME01] N. Davies, K. Cheverst, K. Mitchell, and A. Efrat. Using and determining location in a context-sensitive tour guide. *IEEE Computer*, 34(8):35–41, 2001.

[DCS10] A. Dey, A. Cunningham, and C. Sandor. Evaluating depth perception of photorealistic mixed reality visualizations for occluded objects in outdoor environments. In *Proceedings of the Symposium on Virtual Reality Software and Technology*, pages 211–218, 2010.

[EAH08] B. Erol, E. Antúnez, and J.J. Hull. Hotpaper: multimedia interaction with paper using mobile phones. In *Proceeding of the International Conference on Multimedia*, pages 399–408, 2008.

[ERZHR10] E. Enrico Rukzio, A. Zimmermann, N. Henze, and X. Righetti. Mobile interaction with the real world: Introduction to the special issue. *International Journal of Mobile Human Computer Interaction*, 2(3):i–v, 2010.

[Fit54] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, 47(6), 1954.

[Fit93] G. W. Fitzmaurice. Situated information spaces and spatially aware palmtop computers. *Communications of the ACM*, 36(7):39–49, 1993.

[FMHS93] S. Feiner, B. MacIntyre, M. Haupt, and E. Solomon. Windows on the world: 2D windows for 3D augmented reality. In *Proceedings of the symposium on User interface software and technology*, pages 145–155, 1993.

[FMS92] S. Feiner, B. MacIntyre, and D. Seligmann. Annotating the real world with knowledge-based graphics on a see-through head-mounted display. *Proceedings of the Conference on Graphics Interface*, pages 78–85, 1992.

[FMS93] S. Feiner, B. Macintyre, and D. Seligmann. Knowledge-based augmented reality. *Communications of the ACM*, 36(7):53–62, 1993.

[For08] NFC Forum. Essentials for Successful NFC Mobile Ecosystems. `http://www.nfc-forum.org/resources/white_papers/NFC_Forum_Mobile_NFC_Ecosystem_White_Paper.pdf`, 2008. White paper, NFC Forum.

[FRD$^+$07] O. Falke, E. Rukzio, U. Dietz, P. Holleis, and A. Schmidt. Mobile services for near field communication. *Technical Report of the Ludwig-Maximilians-Universität (LMUMI-2007-1)*, 2007.

[FXL$^+$05] X. Fan, X. Xie, Z. Li, M. Li, and W.Y. Ma. Photo-to-search: using multimodal queries to search the web from mobile devices. In *Proceedings of the international workshop on Multimedia information retrieval*, pages 143–150, 2005.

[FZC93] G. W. Fitzmaurice, S. Zhai, and M. H. Chignell. Virtual reality for palmtop computers. *ACM Transactions on Information Systems (TOIS)*, 11(3):197–218, 1993.

[GAMS08] Q. Gan, J. Attenberg, A. Markowetz, and T. Suel. Analysis of geographic queries in a search engine log. In *Proceedings of the international workshop on Location and the web*, pages 49–56, 2008.

[GAR$^+$09] A. Greaves, P. Akerman, E. Rukzio, K. Cheverst, and J. Hakkila. Exploring user reaction to personal projection when used in shared public places: A formative study. In *Proceedings of the Workshop on Context-Aware Mobile Media and Mobile Social Networks*, 2009.

[GBGI08] S. Gustafson, P. Baudisch, C. Gutwin, and P. Irani. Wedge: clutter-free visualization of off-screen locations. In *Proceedings of the International Conference on Human Factors in Computing Systems*, pages 787–796, 2008.

[GBM08] D. Guinard, O. Baecker, and F. Michahelles. Supporting a mobile lost and found community. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 407–410, 2008.

[GDB07] R. Grasset, A. Dunser, and M. Billinghurst. Human-centered development of an ar handheld display. In *Proceedings of the International Symposium on Mixed and Augmented Reality*, pages 1–4, 2007.

[GRCE06] P. Garner, O. Rashid, P. Coulton, and R. Edwards. The mobile phone as a digital SprayCan. In *Proceedings of the International Conference on Advances in computer entertainment technology*, 2006.

[GSF$^+$07] A. Geven, P. Strassl, B. Ferro, M. Tscheligi, and H. Schwab. Experiencing real-world interaction: results from a NFC user experience field trial. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 234–237, 2007.

[GSJ$^+$07] J. Gao, M. Spasojevic, M. Jacob, V. Setlur, E. Reponen, M. Pulkkinen, P. Schloter, and K. Pulli. Intelligent Visual Matching for Providing Context-Aware Information to Mobile Users. In *Supplemental Proceeding of the International Conference Ubiquitous Computing*, 2007.

[GSMM09] S. L. Ghiron, S. Sposato, C. M. Medaglia, and A. Moroni. NFC Ticketing: a Prototype and Usability test of an NFC-based Virtual Ticketing application. In *Proceedings of the International Workshop on Near Field Communication*, pages 45–50, 2009.

[HB08] N. Henze and S. Boll. Snap and share your photobooks. In *Proceedings of the International Conference on Multimedia*, pages 409–418, 2008.

[HB10a] N. Henze and S. Boll. Designing a CD augmentation for mobile phones. In *Proceedings of the International Conference on Human Factors in Computing Systems (extended abstracts)*, pages 3979–3984, 2010.

[HB10b] N. Henze and S. Boll. Evaluation of an Off-Screen Visualization for Magic Lens and Dynamic Peephole Interfaces. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 191–194, 2010.

[HB10c] N. Henze and S. Boll. Push the study to the app store: Evaluating off-screen visualizations for maps in the android market. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, 2010.

[HB11a] N. Henze and S. Boll. It does not Fitts my data! analysing large amounts of mobile touch data. In *Proceedings of the International Conference on Human-Computer Interaction (INTERACT)*, pages 564–567, 2011.

[HB11b] N. Henze and S. Boll. Release your app on sunday eve: finding the best time to deploy apps. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, pages 581–586, 2011.

[HB11c] N. Henze and S. Boll. Who's that girl? handheld augmented reality for printed photo books. In *Proceedings of the International Conference on Human-computer interaction (INTERACT)*, pages 134–151, 2011.

[HB12] N. Henze and S. Boll. Camera-Based Mobile Interaction Techniques for Physical Objects. In *Adjunct Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 101–106, 2012.

[HBO05a] A. Henrysson, M. Billinghurst, and M. Ollila. Face to face collaborative AR on mobile phones. In *Proceedings of the International symposium on Mixed and Augmented Reality*, pages 80–89, 2005.

[HBO05b] A. Henrysson, M. Billinghurst, and M. Ollila. Virtual object manipulation using a mobile phone. In *Proceedings of the International Conference on Augmented tele-existence*, pages 164–171, 2005.

[HBR$^+$08] N. Henze, G. Broll, E. Rukzio, M. Rohs, and A. Zimmermann. Mobile interaction with the real world. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, 2008.

[HBW10] A. Hang, G. Broll, and A. Wiethoff. Visual design of physical user interfaces for NFC-based mobile interaction. In *Proceedings of the Conference on Designing Interactive Systems*, pages 292–301, 2010.

[Hen12a] N. Henze. Hit it!: an apparatus for upscaling mobile hci studies. In *Proceedings of the International Conference on Human Factors in Computing Systems (extended abstracts)*, pages 1333–1338, 2012.

[Hen12b] N. Henze. Ten male colleagues took part in our lab-study about mobile texting. In *Proceedings of the Workshop on Designing and Evaluating Text Entry Methods*, 2012.

[Hes08] R. Hess. Sift feature detector. `http://web.engr.oregonstate.edu/~hess`, 2008.

[HHB06] N. Henze, W. Heuten, and S. Boll. Non-intrusive somatosensory navigation support for blind pedestrians. In *Proceedings of the European Conference Haptics (EuroHaptics)*, volume 2006, pages 459–464, 2006.

[HHB07a] W. Heuten, N. Henze, and S. Boll. Auditorypong – playing pong in the dark. In *Proceedings of Audio Mostly the Conference on Interaction with Sound*, pages 134–147, 2007.

[HHB07b] W. Heuten, N. Henze, and S. Boll. Interactive exploration of city maps with auditory torches. In *Proceedings of the International Conference on Human Factors in Computing Systems (extended abstracts)*, pages 1959–1964, 2007.

[HHB10] T. Hesselmann, N. Henze, and S. Boll. Flashlight: optical communication between mobile phones and interactive tabletops. In *Proceedings of the International Conference on Interactive Tabletops and Surfaces*, pages 135–138, 2010.

[HHPB08] W. Heuten, N. Henze, M. Pielot, and S. Boll. Tactile wayfinder: a non-visual support system for wayfinding. In *Proceedings of the Nordic Conference on Human-Computer Interaction*, pages 172–181, 2008.

[HLB$^+$10] N. Henze, A. Löcken, S. Boll, T. Hesselmann, and M. Pielot. Free-hand gestures for music playback: deriving gestures with a user-centred process. In *Proceedings of the International Conference on Mobile and Ubiquitous Multimedia*, 2010.

[HLL$^+$07] N. Henze, M. Lim, A. Lorenz, M. Mueller, X. Righetti, E. Rukzio, A. Zimmermann, N. Magnenat-Thalmann, S. Boll, and D. Thalmann. Contextual bookmarks. In *Proceedings of the Workshop on Mobile Interaction with the Real World*, 2007.

[HMB07] A. Henrysson, J. Marshall, and M. Billinghurst. Experiments in 3d interaction for mobile phone ar. In *Proceedings of the International Conference on Computer graphics and interactive techniques in Australia and Southeast Asia*, pages 187–194, 2007.

[HOHS07] P. Holleis, F. Otto, H. Hussmann, and A. Schmidt. Keystroke-level model for advanced mobile phone interaction. In *Proceedings of the International Conference on Human Factors in Computing Systems*, pages 1505–1514, 2007.

[HP12] N. Henze and B. Poppinga. Measuring latency of touch and tactile feedback in touchscreen interaction using a mobile game. In *Proceedings of the International Workshop on Research in the Large*, pages 23–26, 2012.

[HPB10] N. Henze, B. Poppinga, and S. Boll. Experiments in the wild: Public evaluation of off-screen visualizations in the android market. In *Proceedings of the Nordic Conference on Human-Computer Interaction*, pages 675–678, 2010.

[HPP$^+$11] N. Henze, M. Pielot, B. Poppinga, T. Schinke, and S. Boll. My app is an experiment: Experience from user studies in mobile app stores. *International Journal of Mobile Human Computer Interaction (IJMHCI)*, 3(4):71–91, 2011.

[HR08] R. Hardy and E. Rukzio. Touch & interact: touch-based interaction of mobile phones with displays. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 245–254, 2008.

[HRB11] N. Henze, E. Rukzio, and S. Boll. 100,000,000 taps: analysis and improvement of touch performance in the large. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 133–142, 2011.

[HRB12] N. Henze, E. Rukzio, and S. Boll. Observational and experimental investigation of typing behaviour using virtual keyboards on mobile devices. In *Proceedings of the International Conference on Human Factors in Computing Systems*, 2012.

[HRG08] A. Hang, E. Rukzio, and A. Greaves. Projector phone: a study of using mobile phones with integrated projector for interaction with maps. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 207–216, 2008.

[HRHW11] R. Hardy, E. Rukzio, P. Holleis, and M. Wagner. Mystate: sharing social and contextual information through touch interactions with tagged objects. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 475–484, 2011.

[HRL$^+$08] N. Henze, E. Rukzio, A. Lorenz, X. Righetti, and S. Boll. Physical-virtual linkage with contextual bookmarks. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services (extended abstracts)*, 2008.

[HRR$^+$08] N. Henze, R. Reiners, X. Righetti, E. Rukzio, and S. Boll. Services surround you: Physical-virtual linkage with contextual bookmarks. *The Visual Computer*, 24(7), 2008.

[HS88] S. G. Hart and L. E. Staveland. Development of NASA-TLX: Results of empirical and theoretical research. *Human mental workload*, 1:139–183, 1988.

[HSB09] N. Henze, T. Schinke, and S. Boll. What is That? Object Recognition from Natural Features on a Mobile Phone. *Proceedings of the Workshop on Mobile Interaction with the Real World*, 2009.

[HSB11] P. Holleis, M. Scherr, and G. Broll. A revised mobile klm for interaction with multiple nfc-tags. *Proceedings of the International Conference on Human-Computer Interaction (INTERACT)*, pages 204–221, 2011.

[HWI$^+$07] J. Häikiö, A. Wallin, M. Isomursu, H. Ailisto, T. Matinmikko, and T. Huomo. Touch-based user interface for elderly users. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 289–296, 2007.

[Int06a] International Organization for Standardization: Information Technology. Automatic identification and data capture techniques - bar code symbology - qr code. ISO/IEC 18004:2006, 2006.

[Int06b] International Organization for Standardization: Information Technology. Automatic identification and data capture techniques - data matrix bar code symbology specification. ISO/IEC 16022:2006, 2006.

[Int08] Intel Corporation. Open source computer vision library website. `http://www.intel.com/technology/computing/opencv`, 2008.

[IU97] H. Ishii and B. Ullmer. Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proceedings of the International Conference on Human Factors in Computing Systems*, pages 234–241, 1997.

[JW08] I. A. Junglas and R. T. Watson. Location-based services. *Communications of the ACM*, 51(3):65–69, 2008.

[KBM$^+$02] T. Kindberg, J. Barton, J. Morgan, G. Becker, D. Caswell, P. Debaty, G. Gopal, M. Frid, V. Krishnan, H. Morris, J. Schettino, B. Serra, and M. Spasojevic. People, places, things: Web presence for the real world. *Mobile Networks and Applications*, 7(5):365–376, 2002.

[KFH$^+$06] M. Kranz, S. Freund, P. Holleis, A. Schmidt, and H. Arndt. Developing Gestural Input. In *Proceedings of the International Workshop on Smart Appliances and Wearable Computing*, 2006.

[KGWL03] S. R. Klemmer, J. Graham, G. J. Wolff, and J. A. Landay. Books with voices: paper transcripts as a physical interface to oral histories. In *Proceedings of the International Conference on Human Factors in Computing Systems*, pages 89–96, 2003.

[Kin02] T. Kindberg. Implementing physical hyperlinks using ubiquitous identifier resolution. In *Proceedings of the International Conference on World Wide Web*, pages 191–199, 2002.

[KMG+09] S. Karpischek, C. Marforio, M. Godenzi, S. Heuel, and F. Michahelles. Swisspeaks–mobile augmented reality to identify mountains. In *Proceedings of the Workshop on Let's Go Out: Research in Outdoor Mixed and Augmented Reality*, 2009.

[Kön09] W. König. Synchrone und Asynchrone Präsentationstechniken. Bachelor thesis, Media Informatics and Multimedia Systems Group, University of Oldenburg, 2009.

[KOKV06] J. Korhonen, T. Ojala, M. Klemola, and P. Vaananen. mtag-architecture for discovering location specific mobile web services using rfid and its evaluation with two case studies. In *Proceedings of the joint International Conference on Telecommunications and International Conference on Internet and Web Applications and Services*, 2006.

[Küp05] A. Küpper. *Location-based services*. Wiley Online Library, 2005.

[KT07] H. Kato and K.T. Tan. Pervasive 2d barcodes for camera phone applications. *IEEE Pervasive Computing*, 6(4):76–85, 2007.

[Lan11] E. Langbehn. Visualisierung von Off-Screen Objekten in mobiler Mixed Reality. Bachelor thesis, Media Informatics and Multimedia Systems Group, University of Oldenburg, 2011.

[Löc10] A. Löcken. Musical Webcam - Eine alternative Schnittstelle zur Steuerung der Musikwiedergabe. Bachelor thesis, Media Informatics and Multimedia Systems Group, University of Oldenburg, 2010.

[LH99] P. Ljungstrand and L.E. Holmquist. WebStickers: using physical objects as WWW bookmarks. In *Proceedings of the International Conference on Human Factors in Computing Systems (extended abstracts)*, pages 332–333, 1999.

[LHP+11] A. Löcken, T. Hesselmann, M. Pielot, N. Henze, and S. Boll. User-centred process for the definition of free-hand gestures applied to controlling music playback. *Multimedia Systems*, pages 1–17, 2011.

[LLLW10] C. Liao, Q. Liu, B. Liew, and L. Wilcox. Pacer: fine-grained interactive paper via camera-touch hybrid gestures on a cell phone. In *Proceedings of the International Conference on Human Factors in Computing Systems*, pages 2441–2450, 2010.

[LMGY04] T. Liu, A.W. Moore, A. Gray, and K. Yang. An investigation of practical approximate nearest neighbor algorithms. *Advances in neural information processing systems*, 2004.

[Low99] D.G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the International Conference on Computer Vision*, pages 1150–1157, 1999.

[Low04] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[LR06] K. Leichtenstern and E. Rukzio. Mobile interaction in smart environments. In *Proceedings of the International Conference on Pervasive Computing*, pages 43–48, 2006.

[LRH00] P. Ljungstrand, J. Redström, and L.E. Holmquist. WebStickers: using physical tokens to access, manage and share bookmarks to the Web. In *Proceedings of the Conference on Designing augmented reality environments*, pages 23–31, 2000.

[MBGH07] K. Mäkelä, S. Belt, D. Greenblatt, and J. Häkkilä. Mobile interaction with visual and RFID tags: a field study on user perceptions. In *Proceedings of the International Conference on Human Factors in Computing Systems*, pages 991–994, 2007.

[MBMC11] A. Morrison, O. Brown, D. McMillan, and M. Chalmers. Informed consent and users' attitudes to logging in large scale trials. In *Proceedings of the International Conference on Human Factors in Computing Systems (extended abstracts)*, pages 1501–1506, 2011.

[MCUP04] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, 2004.

[MFP$^+$12] H. Müller, J. Fortmann, M. Pielot, T. Hesselmann, B. Poppinga, W. Heuten, N. Henze, and S. Boll. Ambix: Designing ambient light information displays. In *Proceedings of the Workshop on Designing Interactive Lighting*, 2012.

[MLB04] M. Mohring, C. Lessig, and O. Bimber. Video see-through AR on consumer cell-phones. In *Proceedings of the International Symposium on Mixed and Augmented Reality*, pages 252–253, 2004.

[MMB$^+$10] D. McMillan, A. Morrison, O. Brown, M. Hall, and M. Chalmers. Further into the wild: Running worldwide trials of mobile systems. In *Proceedings of the International Conference on Pervasive Computing*, pages 210–227, 2010.

[MML$^+$11] A. Morrison, A. Mulloni, S. Lemmelä, A. Oulasvirta, G. Jacucci, P. Peltonen, D. Schmalstieg, and H. Regenbrecht. Collaborative use of mobile augmented reality with paper maps. *Computers & Graphics*, 35:789–799, 2011.

[MMRGS10] C. Magnusson, M. Molina, K. Rassmus-Gröhn, and D. Szymczak. Pointing for non-visual orientation and navigation. In *Proceedings of the Nordic Conference on Human-Computer Interaction: Extending Boundaries*, pages 735–738, 2010.

[MOP$^+$09] A. Morrison, A. Oulasvirta, P. Peltonen, S. Lemmela, G. Jacucci, G. Reitmayr, J. Näsänen, and A. Juustila. Like bees around the hive: a comparative study of a mobile augmented reality map. In *Proceedings of the International Conference on Human Factors in Computing Systems*, pages 1889–1898, 2009.

[MRGS10] C. Magnusson, K. Rassmus-Gröhn, and D. Szymczak. Scanning angles for directional pointing. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 399–400, 2010.

[MWW06] S. Mehra, P. Werkhoven, and M. Worring. Navigating on handheld displays: Dynamic versus static peephole navigation. *Transactions on Computer-Human Interaction*, 13(4):448–457, 2006.

[Naf08] M. Naffati. Entwicklung eines multimodalen Tangible User Interface. Bachelor thesis, Media Informatics and Multimedia Systems Group, University of Oldenburg, 2008.

[Nee10] C. Needham. Imdb message for 2010. `http://www.imdb.com/czone/top_msg`, 2010.

[Nie94] J. Nielsen. Heuristic evaluation. *Usability inspection methods*, pages 25–62, 1994.

[NIH01] J. Newman, D. Ingram, and A. Hopper. Augmented reality in a wide area sentient environment. In *Proceedings International Symposium on Augmented Reality*, pages 77–86, 2001.

[Nok06] Nokia - Press Release. Nokia launches china's first nfc mobile payment trial in xiamen. http://presse.nokia.fr/2006/06/27/nokia-launches-chinas-first-nfc-mobile-payment-trial-in-xiamen/, 2006.

[NRW06] B. Nath, F. Reynolds, and R. Want. RFID technology and applications. *IEEE Pervasive Computing*, 5(1):22–24, 2006.

[NS06] D. Nister and H. Stewenius. Scalable Recognition with a Vocabulary Tree. In *Proceedings of the International Conference Computer Vision and Pattern Recognition*, pages 2161–2168, 2006.

[Nüs07] D. Nüss. Kontextsensitive visuelle Erstellung von Lesezeichen für Medien. Diploma thesis, Media Informatics and Multimedia Systems Group, University of Oldenburg, 2007.

[OBF94] R. Ohbuchi, M. Bajura, and H. Fuchs. Case study: observing a volume rendered fetus within a pregnant patient. In *Proceedings of the Conference on Visualization*, pages 364–368, 1994.

[OF09] O. Oda and S. Feiner. Interference avoidance in multi-user hand-held augmented reality. In *Proceedings of the International Symposium on Mixed and Augmented Reality*, pages 13–22, 2009.

[OHH04] E. Ohbuchi, H. Hanaizumi, and L.A. Hock. Barcode readers using the camera device in mobile phones. In *Proceedings of the International Conference on Cyberworlds*, pages 260–265, 2004.

[OJ06] S. Ortiz Jr. Is near-field communication close to success? *IEEE Computer*, 39(3):18–20, 2006.

[OS10] P. Ozcan and F. Santos. The market that never was: Clashing frames and failed coalitions in mobile payments. *IESE Research Papers*, 2010.

[Ote99] A. Otero. A robust software barcode reader using the hough transform. In *Proceedings of the International Conference on Information Intelligence and Systems*, page 313, 1999.

[OTRK05] A. Oulasvirta, S. Tamminen, V. Roto, and J. Kuorelahti. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile hci. In *Proceedings of the International Conference on Human Factors in Computing Systems*, pages 919–928, 2005.

[PCAA10] P. Pombinho, M.B. Carmo, A.P. Afonso, and H. Aguiar. Location and orientation based queries on mobile environments. In *Proceedings of the International Conference on Computer Graphics, Visualization, Computer Vision and Image Processing*, 2010.

[PHB08] M. Pielot, N. Henze, and S. Boll. Sensing your social net at night. In *Proceedings of the Workshop on Night and darkness: Interaction after dark*, pages 5–10, 2008.

[PHB09] M. Pielot, N. Henze, and S. Boll. Supporting map-based wayfinding with tactile cues. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 23:1–23:10, 2009.

[PHB11] M. Pielot, N. Henze, and S. Boll. Experiments in app stores – how to ask users for their consent? In *Proceedings of the Workshop on Ethics, logs & videotape*, 2011.

[PHF$^+$12] B. Poppinga, N. Henze, J. Fortmann, W. Heuten, and S. Boll. Ambiglasses-information in the periphery of the visual field. In *Proceedings of Mensch & Computer*, pages 153–162, 2012.

[PHHB07] M. Pielot, N. Henze, W. Heuten, and S. Boll. Tangible user interface for the exploration of auditory city maps. In *Proceedings of the International Workshop on Haptic and Audio Interaction Design*, pages 86–97, 2007.

[PHHB08] M. Pielot, N. Henze, W. Heuten, and S. Boll. Evaluation of continuous direction encoding with tactile belts. In *Proceedings of the International Workshop on Haptic and Audio Interaction Design*, 2008.

[PHN$^+$08] M. Pielot, N. Henze, C. Nickel, C. Menke, S. Samadi, and S. Boll. Evaluation of Camera Phone Based Interaction to Access Information Related to Posters. In *Proceedings of the Workshop on Mobile Interaction with the Real World*, pages 93–103, 2008.

[PHZB12] M. Pielot, W. Heuten, S. Zerhusen, and S. Boll. Dude, where's my car? in-situ evaluation of a tactile car finder. In *Proceedings of the Nordic Conference on Human-Computer Interaction*, 2012.

[Pie07] M. Pielot. Tangible User Interface zur Exploration räumlich sonifizierter Stadtpläne. Diploma thesis, Media Informatics and Multimedia Systems Group, University of Oldenburg, 2007.

[PKB10] M. Pielot, O. Krull, and S. Boll. Where is my team: supporting situation awareness with tactile displays. In *Proceedings of the International Conference on Human Factors in Computing Systems*, pages 1705–1714, 2010.

[Pop07] B. Poppinga. Beschleunigungsbasierte 3D-Gestenerkennung mit dem Wii-Controller. Bachelor thesis, Media Informatics and Multimedia Systems Group, University of Oldenburg, 2007.

[Pop08] B. Poppinga. Entwicklung eines eingebetteten Systems zur Unterstützung von Fahradfahrern durch eine multimodale Benutzungsschnittstelle. Diploma thesis, Media Informatics and Multimedia Systems Group, University of Oldenburg, 2008.

[PPB10] M. Pielot, B. Poppinga, and S. Boll. Pocketnavigator: vibro-tactile waypoint navigation for everyday mobile devices. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 423–426, 2010.

[PPHB10] B. Poppinga, M. Pielot, N. Henze, and S. Boll. Unsupervised user observation in the app store: Experiences with the sensor-based evaluation of a mobile pedestrian navigation application. In *Proceedings of the Workshop on Observing the Mobile User Experience*, pages 41–44, 2010.

[PPHB12] M. Pielot, B. Poppinga, W. Heuten, and S. Boll. Pocketnavigator: Studying tactile navigation systems in-situ. In *Proceedings of the International Conference on Human Factors in Computing Systems*, pages 3131–3140, 2012.

[RA00] J. Rekimoto and Y. Ayatsuka. CyberCode: designing augmented reality environments with visual tags. In *Proceedings of the Conference on Designing augmented reality environments*, pages 1–10, 2000.

[Ram07] G. Ramkumar. Image recognition as a method for opt-in and applications for mobile marketing. *International Journal of Mobile Marketing*, 2(2):42–49, 2007.

[RB10] J. Rico and S. Brewster. Usable gestures for mobile interfaces: evaluating social acceptability. In *Proceedings of the International Conference on Human Factors in Computing Systems*, pages 887–896, 2010.

[RD06] E. Rosten and T. Drummond. Machine learning for high-speed corner detection. In *Proceedings of the European Conference on Computer Vision*, pages 430–443, 2006.

[Rei11] D. Reilly. Reaching the same point: Effects on consistency when pointing at objects in the physical environment without feedback. *International Journal of Human-Computer Studies*, 69(1-2):9–18, 2011.

[Rek97] J. Rekimoto. Pick-and-drop: a direct manipulation technique for multiple computer environments. In *Proceedings of the symposium on User interface software and technology*, pages 31–39, 1997.

[RG04] M. Rohs and B. Gfeller. Using camera-equipped mobile phones for interacting with real-world objects. In *Proceedings of the International Conference on Pervasive Computing*, pages 265–271, 2004.

[RK03] R.E. Rice and J.E. Katz. Comparing internet and mobile phone usage: digital divides of usage, adoption, and dropouts. *Telecommunications Policy*, 27(8-9):597–623, 2003.

[RLC$^+$06] E. Rukzio, K. Leichtenstern, V. Callaghan, P. Holleis, A. Schmidt, and J. Chin. An experimental comparison of physical mobile interaction techniques: Touching, pointing and scanning. In *Proceeding of the International Conference Ubiquitous Computing*, pages 87–104, 2006.

[RLFS07] C. Roduner, M. Langheinrich, C. Floerkemeier, and B. Schwarzentrub. Operating appliances with mobile phones–strengths and limits of a universal interaction device. In *Proceedings of the International Conference on Pervasive Computing*, pages 198–215, 2007.

[RLS07] E. Rukzio, K. Leichtenstern, and A. Schmidt. Mobile interaction with the real world: An evaluation and comparison of physical mobile interaction techniques. In *Proceedings of the European Conference on Ambient Intelligence*, pages 1–18, 2007.

[RN95] J. Rekimoto and K. Nagao. The world through the computer: Computer augmented interaction with real world environments. In *Proceedings of the symposium on User interface and software technology*, 1995.

[RO08] M. Rohs and A. Oulasvirta. Target acquisition with camera phones when used as magic lenses. In *Proceedings of the International Conference on Human Factors in Computing Systems*, pages 1409–1418, 2008.

[Roh05] M. Rohs. Real-world interaction with camera phones. *Proceedings of the International Symposium on Ubiquitous Computing Systems*, pages 74–89, 2005.

[Roh07] M. Rohs. Marker-based embodied interaction for handheld augmented reality games. *Journal of Virtual Reality and Broadcasting*, 4(5), 2007.

[ROS11] M. Rohs, A. Oulasvirta, and T. Suomalainen. Interaction with magic lenses: real-world validation of a fitts' law model. In *Proceedings of the International Conference on Human Factors in Computing Systems*, pages 2725–2728, 2011.

[RPF$^+$06] E. Rukzio, M. Paolucci, T. Finin, P. Wisner, and T. Payne. Mobile interaction with the real world. In *Proceedings of the International Conference on Human-computer interaction with mobile devices and services*, 2006.

[RPKV06] V. Roto, A. Popescu, A. Koivisto, and E. Vartiainen. Minimap: a web page visualization method for mobile phones. In *Proceedings of the International Conference on Human Factors in Computing Systems*, pages 35–44, 2006.

[RSA06] J. Riekki, T. Salminen, and I. Alakärppä. Requesting pervasive services by touching RFID tags. *IEEE Pervasive Computing*, pages 40–46, 2006.

[RSKH07] M. Rohs, J. Schöning, A. Krüger, and B. Hecht. Towards real-time markerless tracking of magic lenses on paper maps. In *Proceedings of the International Conference on Pervasive Computing (Late Breaking Results)*, pages 69–72, 2007.

[RSR$^+$07] M. Rohs, J. Schöning, M. Raubal, G. Essl, and A. Krüger. Map navigation with mobile devices: virtual versus physical movement with and without visual context. In *Proceedings of the International Conference on Multimodal interfaces*, pages 146–153, 2007.

[RSS$^+$09] M. Rohs, R. Schleicher, J. Schöning, G. Essl, A. Naumann, and A. Krüger. Impact of item density on the utility of visual context in magic lens interactions. *Personal and Ubiquitous Computing*, 13(8):633–646, 2009.

[Ruk06] E. Rukzio. *Physical mobile interactions: mobile devices as pervasive mediators for interactions with the real world.* PhD thesis, University of Munich, 2006.

[RZHR09] E. Rukzio, A. Zimmermann, N. Henze, and X. Righetti. Mobile interaction with the real world: Introduction to the special issue. *International Journal of Mobile Human Computer Interaction*, 2009.

[SB09] P. Sandhaus and S. Boll. From usage to annotation: analysis of personal photo albums for semantic photo understanding. In *Proceedings of the Workshop on Social media*, pages 27–34, 2009.

[SCDM10] C. Sandor, A. Cunningham, A. Dey, and V.V. Mattila. An augmented reality x-ray system based on visual saliency. In *Proceedings of the International Symposium on Mixed and Augmented Reality*, pages 27–36, 2010.

[Sch09] T. Schinke. Where do I want to go? Visualisierung von Offscreen-Objekten in Augmented Reality Anwendungen auf mobilen Endgeräten. Diploma thesis, Media Informatics and Multimedia Systems Group, University of Oldenburg, 2009.

[SHB10] T. Schinke, N. Henze, and S. Boll. Visualization of Off-Screen Objects in Mobile Augmented Reality. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 313–316, 2010.

[SJWS02] A. Spink, B.J. Jansen, D. Wolfram, and T. Saracevic. From e-sex to e-commerce: Web search changes. *IEEE Computer*, 35(3):107–109, 2002.

[SMGB11] M. Stroila, J. Mays, B. Gale, and J. Bach. Augmented transit maps. In *Proceedings of the Workshop on Applications of Computer Vision*, pages 485–490, 2011.

[SPHB08] T. Schlömer, B. Poppinga, N. Henze, and S. Boll. Gesture recognition with a wii controller. In *Proceedings of the International Conference on Tangible and embedded interaction*, pages 11–14, 2008.

[SRC05] K.A. Siek, Y. Rogers, and K.H. Connelly. Fat finger worries: How older and younger users physically interact with PDAs. *Proceedings of the International Conference on Human-computer interaction (INTERACT)*, pages 267–280, 2005.

[SRHH09] K. Seewoonauth, E. Rukzio, R. Hardy, and P. Holleis. Touch & connect and touch & select: interacting with a computer by touching it with a mobile phone. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 1–9, 2009.

[SRP08] I. Sánchez, J. Riekki, and M. Pyykknen. Touch & control: Interacting with services by touching RFID tags. In *Proceedings of the International Workshop on RFID Technology*, pages 12–13, 2008.

[SRRP08] I. Sánchez, J. Riekki, J. Rousu, and S. Pirttikangas. Touch & Share: RFID based ubiquitous file containers. In *Proceedings of the International Conference on Mobile and Ubiquitous Multimedia*, pages 57–63, 2008.

[SS04] D. Svanaes and G. Seland. Putting the users center stage: role playing and low-fi prototyping enable end users to design mobile systems. In *Proceedings of the International Conference on Human Factors in Computing Systems*, 2004.

[SSH97] C. Sant'Anselmo, R. Sant'Anselmo, and D. C. Hooper. Identification symbol system and method with orientation mechanism, March 18. 1997. US Patent 5,612,524.

[SSPG11] G. Schall, J. Schöning, V. Paelke, and G. Gartner. A survey on augmented maps and environments: Approaches, interactions and applications. *Advances in Web-based GIS, Mapping Services and Applications*, pages 207–225, 2011.

[ST10] J. Sheridan and J. Tennison. Linking uk government data. In *Proceedings of the Workshop on Linked Data on the Web*, 2010.

[Sut68] I.E. Sutherland. A head-mounted three dimensional display. In *Proceedings of the joint computer conference*, pages 757–764, 1968.

[SV07] J. Schwieren and G. Vossen. Implementing Physical Hyperlinks for Mobile Applications Using RFID Tags. In *Proceedings of the International Symposium on Database Engineering and Applications*, pages 154–162, 2007.

[Syn09] Synovate. Global mobile phone survey shows the mobile is a 'remote control' for life. http://www.synovate.com/news/article/2009/09/global-mobile-phone-survey-shows-the-mobile-is-a-remote-control-for-life.html, 2009.

[TCG$^+$08] G. Takacs, V. Chandrasekhar, N. Gelfand, Y. Xiong, W.C. Chen, T. Bismpigiannis, R. Grzeszczuk, K. Pulli, and B. Girod. Outdoors augmented reality on mobile phone using loxel-based visual feature organization. In *Proceeding of the International Conference on Multimedia information retrieval*, pages 427–434, 2008.

[TF09] Y. Tokusho and S. Feiner. Prototyping an outdoor mobile augmented reality street view application. In *Proceedings of the Workshop on Let's Go Out: Research in Outdoor Mixed and Augmented Reality*, 2009.

[The11] The Wikipedia. Wikipedia, the free encyclopedia, 2011. [Online; accessed 18-May-2011].

[TM08] T. Tuytelaars and K. Mikolajczyk. Local invariant feature detectors: a survey. *Foundations and Trends in Computer Graphics and Vision*, 3(3):177–280, 2008.

[TMN07] K. Tollmar, T. Möller, and B. Nilsved. A picture is worth a thousand keywords: exploring mobile image-based web search. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 421–428, 2007.

[Tog03] B. Tognazzini. First principles of interaction design. *Interaction design solutions for the real world, AskTog*, 2003.

[TSM$^+$07] E. Toye, R. Sharp, A. Madhavapeddy, D. Scott, E. Upton, and A. Blackwell. Interacting with mobile services: an evaluation of camera-phones and visual tags. *Personal and Ubiquitous Computing*, 11(2):97–106, 2007.

[UI00] B. Ullmer and H. Ishii. Emerging frameworks for tangible user interfaces. *IBM systems journal*, 39(3.4):915–931, 2000.

[VK08] E. Veas and E. Kruijff. Vesp'R: design and evaluation of a handheld AR device. In *Proceedings of the International Symposium on Mixed and Augmented Reality*, pages 43–52, 2008.

[VRGMF09] F. Von Reischach, D. Guinard, F. Michahelles, and E. Fleisch. A mobile product recommendation system interacting with tagged products. In *Proceedings of the International Conference on Pervasive Computing and Communications*, pages 1–6, 2009.

[vRM08] F. von Reischach and F. Michahelles. Apriori: A ubiquitous product rating system. In *Proceedings of the Workshop on Pervasive Mobile Interaction Devices*, 2008.

[VT05] P. Välkkynen and T. Tuomisto. Physical browsing research. In *Proceedings of the Workshop on Pervasive Mobile Interaction Devices*, pages 35–38, 2005.

[VTK06] P. Välkkynen, T. Tuomisto, and I. Korhonen. Suggestions for visualising physical hyperlinks. In *Proceedings of the Workshop on Pervasive Mobile Interaction Devices*, pages 245–254, 2006.

[WDH07] J. Wither, S. DiVerdi, and T. Hollerer. Evaluating display types for ar selection and annotation. In *Proceedings of the International Symposium on Mixed and Augmented Reality*, pages 1–4, 2007.

[Wei02] M. Weiser. Hot topics-ubiquitous computing. *IEEE Computer*, 26(10):71–72, 2002.

[WF09] S. White and S. Feiner. Sitelens: situated visualization techniques for urban site visits. In *Proceedings of the International Conference on Human Factors in Computing Systems*, pages 1117–1120, 2009.

[WFGH99] R. Want, K. P. Fishkin, A. Gujar, and B. L. Harrison. Bridging physical and virtual worlds with electronic tags. In *Proceedings of the International Conference on Human Factors in Computing Systems*, pages 370–377, 1999.

[WPLS05] D. Wagner, T. Pintaric, F. Ledermann, and D. Schmalstieg. Towards massively multi-user augmented reality on handheld devices. *Proceedings of the International Conference on Pervasive Computing*, pages 208–219, 2005.

[WRM$^+$08] D. Wagner, G. Reitmayr, A. Mulloni, T. Drummond, and D. Schmalstieg. Pose tracking from natural features on mobile phones. In *Proceedings of the International Symposium on Mixed and Augmented Reality*, pages 125–134, 2008.

[WS03] D. Wagner and D. Schmalstieg. First steps towards handheld augmented reality. In *Proceedings of the International Symposium on Wearable Computers*, pages 127–135, 2003.

[WS09] D. Wagner and D. Schmalstieg. History and Future of Tracking for Mobile Phone Augmented Reality. In *Proceedings of the International Symposiumon Ubiquitous Virtual Reality*, pages 7–10, 2009.

[WSB09] D. Wagner, D. Schmalstieg, and H. Bischof. Multiple target detection and tracking with guaranteed framerates on mobile phones. In *Proceedings of the International Symposium on Mixed and Augmented Reality*, pages 57–64, 2009.

[Yee03] K.P. Yee. Peephole displays: pen interaction on spatially aware handheld computers. In *Proceedings of the International Conference on Human Factors in Computing Systems*, pages 1–8, 2003.

[YSY$^+$09] S. Yoon, J. Seo, J. Yoon, S. Shin, and T.D. Han. A Usability Evaluation of Public Icon Interface. *Proceedings of the International Conference on Human-Computer Interaction (INTERACT)*, pages 540–546, 2009.

[ZHRR09] A. Zimmermann, N. Henze, X. Righetti, and E. Rukzio. Mobile interaction with the real world. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 106–109, 2009.

[ZKG$^+$09] S. Zhai, P.O. Kristensson, P. Gong, M. Greiner, S.A. Peng, L.M. Liu, and A. Dunnigan. Shapewriter on the iPhone: from the laboratory to the real world. In *Proceedings of the International Conference on Human Factors in Computing Systems (extended abstracts)*, pages 2667–2670, 2009.

[ZMG$^+$03] P.T. Zellweger, J.D. Mackinlay, L. Good, M. Stefik, and P. Baudisch. City lights: contextual views in minimal space. In *Proceedings of the International Conference on Human Factors in Computing Systems (extended abstracts)*, pages 838–839, 2003.

[Zmi05] A. Zmijewska. Evaluating wireless technologies in mobile payments-a customer centric approach. In *Proceedings of the International Conference on Mobile Business*, pages 354–362, 2005.