Institute for Visualization and Interactive Systems

University of Stuttgart
Universitätsstraße 38
D–70569 Stuttgart

Bachelorarbeit Nr. 91

The Last-Seen Image: an Image-Based Approach for Finding Lost Objects Using a Head-Mounted Display

Robin Boldt

Course of Study: Softwaretechnik

Examiner: Prof. Dr. Albrecht Schmidt

Supervisor: Dipl.-Inf. Markus Funk,
Dipl.-Inf. Bastian Pfleging

Commenced: November 4, 2013

Completed: May 6, 2014

CR-Classification: H.5.2

Abstract

Considering the current development of commercial head-mounted displays (HMD), it is likely that HMDs will be widely used in the near future. Therefore, it becomes feasible to build systems that rely on HMD technology. Real-world search engines aim at supporting the user's search capabilities in the real world. HMDs are a possible device to guide a user towards the location of an object. To ensure a high usability, it is essential to find suitable location representations of search results on an HMD. Based on previous findings, we present a novel location representation called "last-seen image" to locate objects in a known environment (e.g. the user's home or office building). The last-seen image shows the picture of a sought object including the surrounding context of the object. We implemented a prototype on an HMD using WiFi indoor positioning to provide the proposed visualization as well as a map visualization. We conducted a user study comparing our proposed last-seen image approach to a map based approach. The last-seen image showed to be significantly faster for finding harder hidden objects compared to the map representation. However, the map was favored for finding the correct room. Therefore, we propose a hybrid system using the map representation to find the correct room and using the last-seen image to find the object on room-level.

Kurzfassung

Aufgrund der aktuellen Entwicklung von Head-Mounted Displays (HMD) ist es sehr wahrscheinlich, dass HMDs in Zukunft allgegenwärtig sein werden. Deshalb wird ein System, welches auf der Benutzung von HMDs basiert, realisierbar. Real-World Search Engines unterstützen einen User, verlorene Gegenstände wiederzufinden. HMDs eignen sich zur Repräsentation solcher Suchergebnisse. Um die Benutzbarkeit solcher Systeme zu garantieren, ist es wichtig, eine passende Repräsentationsart zu finden. Aufgrund vorheriger Ergebnisse stellen wir eine neuartige Repräsentationsart zur Objektlokalisierung vor: das Last-Seen Image. Das Last-Seen Image zeigt das Bild eines gesuchten Gegenstandes, welches nicht nur den Gegenstand selbst, sondern auch die Umgebung zeigt. Wir haben einen Prototypen entwickelt, welcher auf einem HMD eine Kartenansicht und das Last-Seen Image bereitstellt. Daraufhin haben wir eine Benutzerstudie durchgeführt, um das Last-Seen Image mit der Kartendarstellung zu vergleichen. Es hat sich gezeigt, dass schwer versteckte Objekte mithilfe des Last-Seen Image deutlich schneller gefunden werden als mit der Kartendarstellung. Jedoch wurde die Karte bevorzugt, um den richtigen Raum zu finden. Deshalb empfehlen wir die Benutzung eines Hybriden Systems, welches die Kartendarstellung verwendet, um den richtigen Raum zu finden. Sobald man sich in dem richtigen Raum befindet wird das Last-Seen Image angezeigt.

Contents

List of Figures

1 Introduction

In 2014 a lot of people are eagerly awaiting the release of Google Glass[1]. Google Glass is a wearable head mounted display (HMD). The release of Google Glass creates a lot of discussion among data protection specialists, users, and the media. For the last decades HMDs have been used almost exclusively in research, aviation, and other highly specialized fields, e.g. in cockpits of modern helicopters. Due to the commercial uprising of HMDs a whole new field of application scenarios is created.

Using an HMD offers numerous benefits compared to using a regular smartphone. The user's hands are free while using the HMD, since the display is mounted in front a user's field of sight. Particularly due to the fact that a lot of HMDs offer interaction based on speech recognition or gestures. Tasks like navigation are improved, since the display is always in the field of sight. Navigation in general demands the attention of the user to be on the street instead of being on the display. Using an HMD a user is able to simultaneously keep an eye on the street and still see the navigation instructions.

Augmented Reality (AR) describes the technology to superimpose an image on a users view of the real world. Today, AR can be found in a lot of different ways. In Navigation, AR applications typical highlight the path to take [Kim and Dey, 2009] [Narzt et al., 2006] or show waypoints to follow [Reitmayr and Schmalstieg, 2004]. There are a lot of AR applications available for Google Glass, e.g. "watchmetalk"[2] or "Mobile Map Tools"[3]. A major business use case using Google Glass describes the use of an HMD to show construction plans [4]. For instance a technician is supported by an HMD to assemble or disassemble some component. The HMD shows the corresponding construction plans by highlighting the process step by step or even showing a video [5]. Every activity that requires the hands can potentially benefit from HMDs.

Considering the current development of commercial head-mounted displays (HMD), it is likely that HMDs become ubiquitous. Not only Google Glass is about to enter the market. Meta[6] for example offers a stereoscopic HMD fit inside an unobtrusive looking sunglasses. However,

[1] http://www.google.de/glass/start/
[2] http://getwatchmetalk.com/
[3] http://glasses.mobilemaptools.com/
[4] http://www.nytimes.com/2014/04/08/technology/google-begins-a-push-to-take-glass-to-work.html
[5] http://www.wearabletechworld.com/topics/wearable-tech/articles/367660-industrial-public-safety-applications-abound-google-glass.htm
[6] https://www.spaceglasses.com/

there are a lot of competitors offering or developing different kinds of HMDs[7]. With the idea of commercial HMDs becoming ubiquitous, systems that entirely base on the usage of HMDs becomes feasible.

Throughout this thesis we describe a system that enables one to find lost items. Due to the fact that losing an object is a time consuming and annoying matter, people try to prevent losing items. Probably everybody experienced the loss of a needed item. There are techniques that can reduce the number of lost items due to improved organization[8] of the environment and one's personal habits. The most common way to improve organization is to place items according to their usage [Kirsh, 1995]. For instance people place their keys close to the exit of the apartment to be able to pick them up when leaving.

We consulted 46 people on how often they are searching for objects that are not recovered immediately. Almost everybody (89%) searches for an object at least once per week for an object. About half of the interviewees (52%) searches for an object on a daily base. Recent research shows that an average person spends 10 to 15 minutes per day searching for lost or misplaced items [Esure, 2012]. According to this study 51% of the participants are claiming to constantly search for lost or misplaced items.

Items seem to get lost even if we are trying not to lose them. A definition of lost items could be: "things that are not where they are supposed to be" [Peters et al., 2004]. Thus, a system representing the current location of each object available solves the issue of losing items.

There are solutions to find and locate people and places but there is no real solution to find items or objects in general. People can be found on Facebook[9], on Google+[10], or with the help of Find My Friends![11]. While Facebook is more a check-in system that allows posting your current location, Google+ and Find My Friends! are real-time tracking systems. Applications like Google Maps[12], Yelp[13] or Foursquare[14], enable a user to search for places in the area like restaurants, bars, and so on. There are systems that can retrieve the place of a photo taken of a landmark [Amato et al., 2010]. Similar technology is available for objects with Amazon Flow[15] or Google Googles[16]. Both apps return information according to the recognized items in the picture.

[7] http://www.glassappsource.com/google-glass/google-glass-competitors.html
[8] also see the National Association of Professional Organizers http://www.napo.net/
[9] https://www.facebook.com/
[10] https://plus.google.com/
[11] https://www.apple.com/apps/find-my-friends/
[12] https://www.google.com/maps/
[13] http://www.yelp.com/
[14] https://foursquare.com/
[15] http://www.amazon.com/A9-Innovations-LLC-Powered-Amazon/dp/B008G318PE
[16] https://play.google.com/store/apps/details?id=com.google.android.apps.unveil

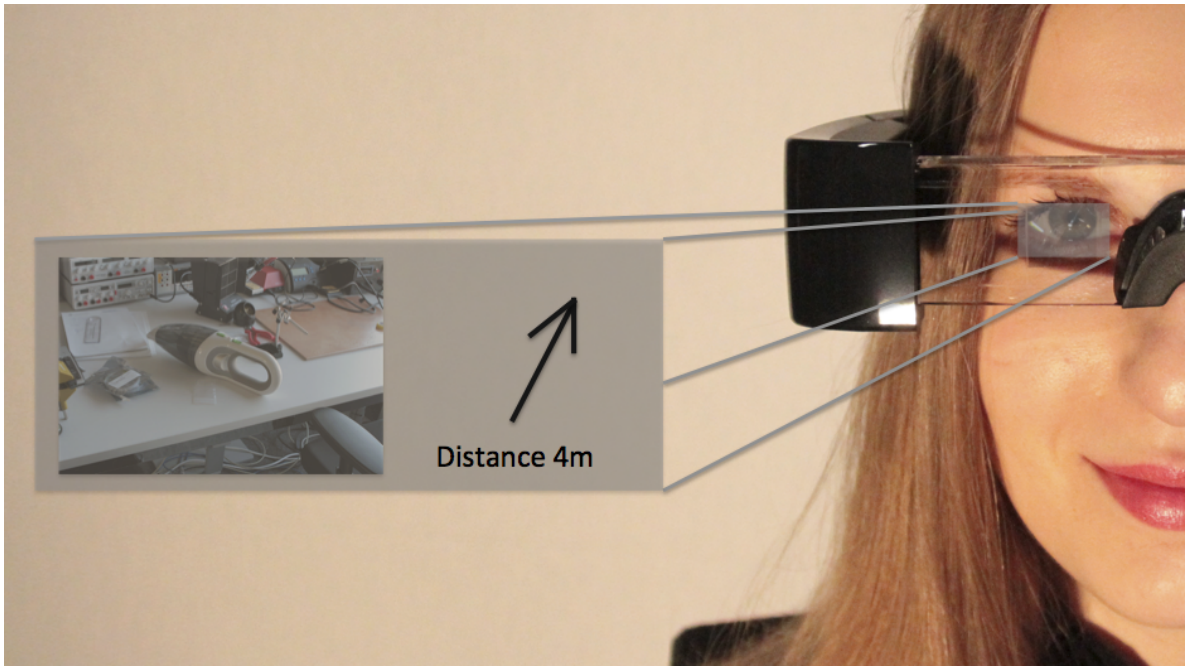

Figure 1.1: A user wearing the prototype. The direction, distance and last-seen image of a sought image is displayed by the prototype.

The current commercial trend in lost item retrieval are key-finders. The are an electronic device that can be attached to your keys or any vital item. Instead of searching for an item in the traditional way, key-finders can support the search task. They can be activated by different techniques depending on the implementation. Some use Bluetooth while others respond to whistling. Once enabled, key-finders drag the user's attention by some sort of signal e.g. emitting an audible signal. A commercial key-finder using Bluetooth technology starts at around 20$. Equipping a lot of items with such tags is highly uneconomic. Not only due to the acquisition cost but due to maintenance costs as well. Additionally, not all items can be easily equipped with Bluetooth tags.

The contribution of this thesis is threefold. (1) We introduce an image based representation (last-seen image) for locating objects in indoor environments (see Figure 1.1). (2) The concept and design space of a real-world search engine using visual object recognition has been explored. (3) We conducted a user study throughout a whole building consisting of several floors, comparing the last-seen image representation to a map based approach.

(1) The last-seen image is a novel object location representation based on the assumption that the environmental context given in an image is sufficient to find an object. The basic suitability of the last-seen image for object retrieval has already been shown in a preliminary study [Boldt et al., 2013]. The challenge throughout this thesis is to show or refute the usability of the last-seen image in a whole building.

(2) The here described system uses wearable HMDs with an integrated camera to recognize objects using visual object recognition. As the system uses camera streams to recognize objects the last-seen image is created as a byproduct. Real-world search engines using visual object recognition have been already proved to be suitable [Funk et al., 2013]. As the design space of real-world search engines has not been explored in depth, we extend the already existing design space.

(3) The last-seen image representation has not been proven to work on building-level. Thus, we conduct a user study throughout the hciLab[17]. The study uses a prototype that was implemented in this project.

If the last-seen image proves to be a suitable object location representation, this will also be a hint towards using a real-world search engine based on visual object recognition. As the last-seen image and a real-world search engine based on visual object recognition benefit mutually from each other. Using the last-seen image as object location representation implies the addition of cameras to the environment. Accordingly, a real-world search engine based on visual object recognition implies the usage of the las-seen image, since the last-seen image is created as a byproduct. However, the last implication is only true if the last-seen image is shown to be the favorite object location representation. Otherwise, the last implication has to be invalidated.

1.1 Structure of the Thesis

The structure of this thesis is described in the following paragraph. We start by showing related work (starting on page 13). The related work covers the identification of objects, indoor positioning systems, and visualization of location. Hereinafter we describe the related work we describe the concept and design space (starting on page 21). The concept provides an in-depth description of the term “real-world search engine”. The design space of a real-world search engine is explored at the end of the concept chapter. As constantly recording images can be highly privacy invasive, we provide an in depth discussion about how to protect the privacy as far as possible. The system chapter (starting on page 33), describes a system that is able to show the representations and navigate a user towards a sought object using an HMD. The experimental evaluation of the system is reported in the study chapter (starting on page 43). At the end of the thesis we give a conclusion (starting on page 53). The concept includes future work and limitations of our proposed system.

[17] http://www.hcilab.org/

1.2 Publication at the Augmented Human 2014

Parts of this thesis have been published on the “5th Augmented Human International Conference” in Kobe, Japan [Funk et al., 2014].

> Markus Funk, Robin Boldt, Bastian Pfleging, Max Pfeiffer, Niels Henze, and Albrecht Schmidt. 2014. Representing indoor location of objects on wearable computers with head-mounted displays. In Proceedings of the 5th Augmented Human International Conference (AH ’14). ACM, New York, NY, USA, , Article 18 , 4 pages. DOI=10.1145/2582051.2582069 http://doi.acm.org/10.1145/2582051.2582069

2 Related Work

This project comprises different topics. Therefore, we want to provide the related work divided in different topics to keep it lucid. A real-world search engine can be divided into three parts: the identification (1) and localization (2) of objects, and the representation of their location (3) to the user. We start describing, how existing systems identify objects and keep track of them. As the major use case of a real-world search engine is indoors, we outline existing indoor positioning systems. The related work ends with showing ways to visualize location, comprising outdoor navigation, indoor navigation, and object localization on room-level.

2.1 Object Identification

The following section shows how existing systems identify objects. This is the most important part of a real-world search engine. The architecture of the entire system is determined by the object identification procedure. Each object identification mechanism implies distinct non functional requirements like: scalability, privacy, reliability, or modifiability. Most of the systems rely on a particular technology. Therefore, we decided to present this section divided into the used technology. The covered technologies are: Bluetooth, radio-frequency identification (RFID), visual markers, and visual feature detection.

Bluetooth is a widely used technology, which almost every smartphone, a lot of laptops, and technical devices are already equipped with. In “Where’s my Stuff” [Kientz et al., 2006] a system working with Bluetooth tags is described. Important objects are equipped with a tag. The tag starts to emit audible signals, if a user is searching for an item. One of the main problems of this solution is the range of Bluetooth signals. Therefore, an extension to the above described system is shown in “Objects calling Home” [Frank et al., 2007]. A network of mobile phones is created. Mobile phones share their objects in range in the created network. Consequently, one can query for an object in the range of another mobile phone.

A lot of commercial systems using Bluetooth have been released recently. Most of them add Bluetooth tags to important items. The current commercial trend in lost item retrieval are so called key-finders. As described previously, key-finders are regular Bluetooth tags added to important items. Even though there are key-finders using other technologies than Bluetooth,

Bluetooth key-finders are the most common ones. The Cobra Tag[1] utilizes Bluetooth to find tagged items using a smartphone. However, things work the other way around as well, a button attached to the tag can be pressed and the smartphone starts to emit an audible signal. Sought items can emit an audible signal as well, as the attached tag is equipped with a speaker. Another commercial system based on Bluetooth is Bringrr[2]. Bluetooth tags are attached to important items and can be searched for by an app. However, the clue about Bringrr is a car charger that notifies the user if an object is missing before driving away.

Today, RFID technology is ubiquitous in our daily life, e.g. in access management, tracking of goods, toll collection, smartdust [Kahn et al., 1999], and contactless payment. Therefore, researchers had the idea to use the already existing environment and extend it. Unfortunately, reality shows that the use of RFID implies huge modification to the environment. A system using RFID technology is practicable for one or two rooms but most probably not for a whole building. Our concerns are especially caused by the unscalability of such systems. Systems using RFID technology can be divided into systems that use mobile RFID readers and those using static RFID readers.

We start by describing systems that use mobile RFID readers. “MagicTouch” [Pederson, 2001] describes a RFID based object identification via wearable RFID readers attached to a users hand. If a user touches a RFID tagged artifact, the reader is able to identify the object and send the location to a backend. This is based on the trivial, nevertheless powerful, assumption that objects are only moved by humans and therefore every movement can be tracked using that system. Instead of using humans, “IteMinder” [Komatsuzaki et al., 2011a] uses an autonomous robot that can automatically move around the room. If the robot finds a tag, it uploads its position to a database.

In contrast to the above, the “LANDMARC-System” [Ni et al., 2004] uses static RFID readers. Trackable objects are equipped with active RFID tags. The idea is to add additional RFID tags named “landmarcs”. A “landmarc” provides a reference location to increase the accuracy of an object’s location. The “MAX-System” [Yap et al., 2008], extends that idea and proposes a hierarchical approach. Each room is equipped with a RFID antenna referred to as base station. Each base station can consist of several sub stations, mostly static objects like furniture. At the lowest level of the hierarchy there are RFID tags that are attached to movable objects like books. The user is enabled to search for objects using query terminals. The terminals returns user specific results according to a sophisticated privacy system. The position of a sought object is stored in the same way as proposed by the hierarchy: base station, sub station, and tag. Therefore, the results are returned in textual form, e.g. “my keys are on my desk”.

A combination of several technologies is proposed in “Find my Stuff” [Nickels et al., 2013]. In general “Find my Stuff” uses the approach of the “MAX-System” for location representation

[1] http://cobraphonetag.com/
[2] http://www.bringrr.com/

(e.g. “my keys are on my desk”), but in contrast “Find my Stuff” equips furniture with RFID readers and ZigBee modules. ZigBee is a technology based on the IEEE 802.15.4 standard. The IEEE 802.15.4 standard describes low-rate wireless personal area networks which focus on ubiquitous communication between devices. “Find my Stuff” provides a robust and calibration free environment, but it is complex to setup, because furniture needs to be equipped with a module setup, as well as objects need to be equipped with tags.

Most previously described systems require the modification of searchable objects as well as the modification of the environment. However, all of them require at least a modification of either one. For instance equipping searchable objects with RFID tags or adding RFID readers to furniture. This is a huge implication to the environment. Imagine an office environment with hundreds of employees and every object needs to be tagged to be searchable. In our opinion, these systems are not practicable because the effort of setting them up does not justify the benefits they offer. If at some future point almost all items are already equipped with RFID tags, these systems will become feasible.

Objects can be detected by their visual features using either markers or object recognition. We start describing approaches using markers. In general markers are adhesive stickers with a printed barcode or a Quick Response Code (QR code) on it.

In “Searchlight” [Butz et al., 2004] objects are equipped with optical markers. The room itself is equipped with a steerable camera and projector unit. Objects can be scanned by the camera and sought objects are highlighted using the projector. Therefore, only objects that are in the sight of the camera can be scanned. Due to the fact that the camera unit is mounted at a fixed position, some objects are never indexed at all, because they are out of the camera’s sight. However, this system was designed for finding books in a shelf which is achieved using the system. We especially want to highlight the fact that “Searchlight” does not depend on an environmental model. In “WebStickers” [Ljungstrand et al., 2000] a system is described that uses barcode readers and adhesive stickers with barcodes to identify objects. Each barcode leads to a web page. Hence, each “WebSticker” can be seen as a bookmark. A camera readable barcode is presented in “CyberCode” [Rekimoto and Ayatsuka, 2000]. “CyberCode” does not only recognize objects, but also retrieve it’s 3D location in the picture, using computer vision technology. In “DrawerFinder” [Komatsuzaki et al., 2011b] a camera is mounted above a shelf. Optical markers are attached to each drawer and will be scanned when the drawer is opened. As soon as a drawer is identified by a visual marker, a picture of its content is taken and will be uploaded. A user can browse through the different drawers and search for items. The system, however, does not allow search queries for a certain object. Therefore, the user needs to browse the images taken of each drawer until the sought object is found.

“Antonius” takes a step forward [Funk et al., 2013]. The authors present a prototype of a real-world search engine that detects objects of the physical world based on their visual appearance. “Antonius” uses a wearable camera system to detect objects and their location.

The location of the wearable prototype is determined using an OptiTrack[3] system. The location of a found object is assumed to be 40 cm in front and 45 cm below the wearable prototype. This assumption was determined experimentally. Users can search for objects and will be shown their location in a 3D-model.

The idea of a wearable visual object detection device is one of the keynotes of the research presented in this thesis. Therefore, no objects need to be marked or tagged in any way. Every system presented in this paper except for "Antonius" requires marking or equipping every searchable object with identifiers. "Antonius" uses an OptiTrack system that is not naturally available in every environment. Thus, an adequate and non intrusive indoor location system has to be found. The following section gives an overview over the existing indoor positioning systems.

2.2 Indoor Positioning

The main use case for a real-world search engine takes place in indoor environments. Due to the fact that the Global Positioning System (GPS) does not work properly in indoor environments, we are in the need of an adequate indoor positioning technology. Our main objective is to provide at least room level accuracy as well as no major changes that need to be done to the environment. In that manner we can provide a scalable and modifiable system.

Over the years several indoor location techniques have been proposed. We start by describing some early systems. On of the early systems is the "Active Badge" [Want et al., 1992]. A user is equipped with a badge that transmits a unique infrared signal every 10 seconds. Each room needs to be equipped with a sensor that receives the signals and therefore can determine the position of the badge. The "Active Bat" [Harter et al., 1999] was developed as an improvement to the "Active Badge". It allows a 3D positioning on centimeter level due to trilateration of ultrasonic receivers mounted at the ceiling. The ultrasonic sound is emitted by a badge and the system measures the times-of-flight. Similar to the "Active Bat", the "Cricket Location Support System" [Priyantha et al., 2000] uses ultrasonic for indoor positioning as well. "RADAR" [Bahl and Padmanabhan, 2000] measures the signal strength of radio frequencies of multiple base stations. The location is calculated using triangulation of the signal strengths and estimates the location within a few meters. With "Smart Floor" [Orr and Abowd, 2000] the user does not need to carry any device which makes the difference to the systems described beforehand. People can be tracked based on their footstep force profile measured by sensors attached to the ground. A 93% accuracy could be achieved using this approach.

[3] http://www.naturalpoint.com/optitrack/

In the following passage we describe current approaches to indoor location. The “RF doormat system” [Ranjan et al., 2013] equips a user with a RF ankle bracelet that is tracked when crossing a doorway. That way a users’ room location can be tracked with an accuracy of 98%. Research on the field of WiFi indoor location [Woodman and Harle, 2008], shows that throughout a building of $8725m^2$ a user can be tracked with an accuracy of 0.73m in 95% of the time. This accuracy is not only achieved by using WiFi fingerprints, but also by using acceleration sensors. Visual position detection by scanning markers can be done discrete [Mulloni et al., 2009], as well as continuous [Kalkusch et al., 2002]. Actual visual indoor location based on feature matching is possible as well [Schroth et al., 2011]. A previously recorded video that is tagged with geographic information is compared to visual input and a location can be calculated accordingly.

Retrieving the indoor location by measuring disturbances of the Earth’s magnetic field, caused by structural steel elements, in a building is described in [Chung et al., 2011]. Before a location can be retrieved the whole building needs to be measured. The disturbances are measured with an array of electronic compasses. An accuracy better than 1m could be in shown 88% of the time across several buildings and floors.

Indoor positioning techniques have been explored throughout the last decades. There are positioning systems for almost all use cases. Sub centimeter accuracy can be achieved with the downside of high costs and major changes to the environment. The OptiTrack system, for instance, can resolve a position with millimeter accuracy. Sub meter accuracy can be achieved using WiFi indoor location. Multiple WiFi access points are available at almost every building. Thus, WiFi indoor location can be done without major changes to the environment or having high costs.

2.3 Visualizing Location

Visualizing a location to make it understandable to a user is a widely researched topic. A lot of work has been done on the field of navigation, especially the outdoor navigation, since it has been more feasible over the last decades by using GPS. We use a hierarchical approach to describe the related work. We start with outdoor visualization, proceed with indoor visualization, and finish with object localization on room-level.

A comparison of textual, map, arrow, and 3D navigation [Kray et al., 2003] concludes without pointing out a favorite route representation. But it was stated that the user’s preferences and abilities are vital factors for selecting a way of route instruction. These representations can be seen as the traditional representations in navigation. Route instructions have been passed in textual or audible way ever since, like “turn right onto Evergreen Terrace”. Not only textual instructions but also maps have been used to navigate for hundreds of years. In contrast the arrow and 3D navigation are new. These route instructions have been introduced with the evolution of computers. A commonly proposed alternative to these representations is a

navigation, based on waypoints to be followed, as described in [Reitmayr and Schmalstieg, 2004]. The user is wearing a head mounted display (HMD) and is shown the waypoints with the use of Augmented Reality (AR). The idea of a waypoint based AR navigation was compared with a map based approach for navigating users using a mobile device [Walther-Franks and Malaka, 2008]. The authors conclude that in some cases the waypoint approach is even better than the map approach. In "Halo" [Baudisch and Rosenholtz, 2003] a visualization technique is described that makes off-screen locations visible to a user. Small displays make it hard to zoom in a map to determine the correct path and still be able to see the destination on the screen. Therefore the interface uses red semicircles at the edge of the screen if the location is outside the screen range. The radius of the semicircle is used to determine the distance to the location. A big radius is used for locations that are far away and vice versa.

Combinations of indoor and outdoor navigation are possible as well. With the "BMW Personal Navigator" [Krüger et al., 2004] a system that combines 3D and map is proposed. The "BMW Personal Navigator" provides navigation in the car as well as pedestrian navigation. With the aid of infrared bacons attached inside a building the user is enabled to navigate inside that buildings . "Drishti" [Ran et al., 2004] is a system supporting blind people in finding their way through known areas as well as unknown areas. The blind receives audible signals to avoid possible obstacles and support way finding. The outdoor positioning is done using Differential GPS. Whereas the indoor positioning is using a ultrasound positioning system. Google Maps[4] is offering outdoor as well as indoor navigation an positioning using maps. At the moment there are more than 10.000 indoor maps available[5] and the number is growing.

Indoor navigation can be handled like outdoor navigation. For example "Footpath" [Link et al., 2011] proposes a map based approach for indoor navigation. But it seems like researchers are taking new approaches. Indoor navigation offers distinct possibilities from outdoor navigation, due to closer distances of junctions, compared with outdoors. Regular streets in a city are several meters apart, in contrary doors inside a building are often less than 1 meter apart. Hence, systems that switch representations based on the accuracy of the position and the contextual information were developed. Accordingly a hybrid system [Butz et al., 2001] that is able to switch between different 2D maps and 3D visualizations has been proposed. Alternatively, a combination of different AR and map representations for indoor navigation is proposed [Mulloni et al., 2011]. The representations are chosen according to the activity of the user as well as the accuracy of the position. As an additional contextual feature info points are introduced to further support the user by switching to an overview map, when the user reaches an info point.

The most important part of finding an object happens on room-level. Different visualizations on a smartphone, with the intend to raise the attention of a user towards an object, have been experimentally compared with each other [Möller et al., 2014]. The authors introduce four

[4]https://www.google.com/maps/
[5]https://maps.google.com/help/maps/indoormaps/faqs.html last accessed May 2014

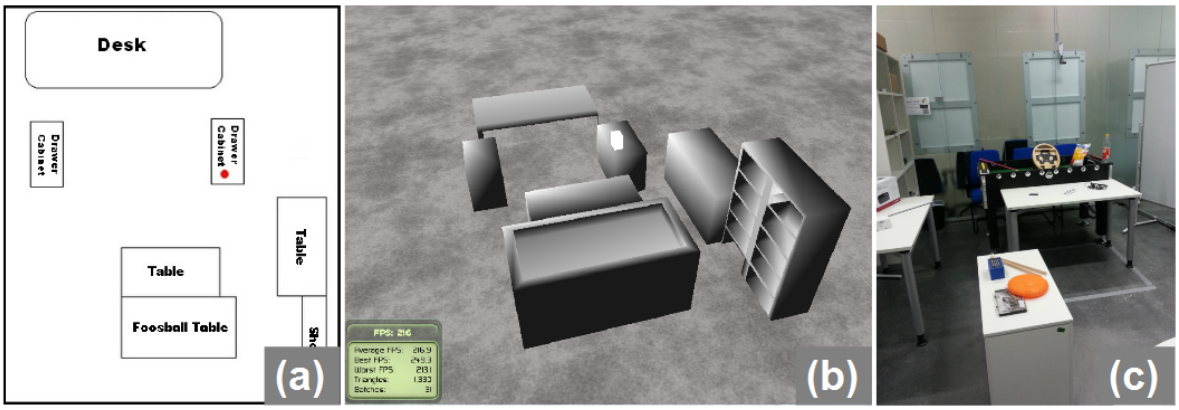

Figure 2.1: A screenshot showing three of the four location representations used in a preliminary study. (a) The map representation with the sought object denoted as a red dot. (b) The 3D representation using the OGRE engine and denoting the position of the sought object as a white rectangular solid. (c) The last-seen image representation of a flying disc.

indicator types: text hint, blur, color scale, and spirit level. They are supposed to motivate a user to aim the smart phone towards an object. If the object is in range of the camera it will be highlighted in the viewport, e.g. a frame will be painted around the object. In Attention Funnel [Biocca et al., 2006] users are equipped with an HMD. The Attention Funnel is an AR technique to guide the attention of a user to any object, person, or place in space. A funnel of rectangles is build that ends at the sought object.

The foundation of this thesis is a preliminary study in which object representations were compared on room-level [Boldt et al., 2013]. Textual, 2D, and 3D representations were compared to an image-based representation in a lab study (see Figure 2.1). The impact each representation had on cognitive load, task completion time, and user experience was assessed. The image-based representation referred to as last-seen image, which includes the object as well as environmental context is shown to be appropriate for object retrieval on room-level. Even though, the lab study could not reveal a statistically significant favorite representation, the authors make statements about the feasibility of each representation. The most valuable asset of the last-seen image would thus be its independence of any environmental models that are mandatory for any of the other representations. Since the preliminary study is the foundation for this thesis each representation is shown in detail.

The **textual** representation is based on nested zones. It returns the location representation with the help of these zones. Zones are used like landmarks [Ni et al., 2004]. Therefore, a textual representation looks like this: Building No. 101 -> Kitchen -> Table -> sunglasses (the sunglasses are on the kitchen table in Building No. 101).

The **3D** representation offers a detailed model of the environment. The sought object will be highlighted in that model to make it visible to the user. The user is enabled to move around in the 3D representation as being in the physical world. However, all physical laws can be violated in the virtual world. Thus, the user will be more powerful compared to a manual search, especially because sought objects can be highlighted through walls and other objects.

The **map** representation shows a map on which sought objects are denoted as a red dot on the map. The map offers a great overview over the situation and enables the user to plan a route towards the objects. The map can even be improved by representing the current location of the user. This could be achieved by denoting the position of the user with another color than the sought object.

The **last-seen image** displays the last taken picture in which the object was recognized by the system. The image does not only show the object, but also the surrounding context of the object like other known objects or furniture.

The field of route representation has been deeply analyzed. There are lots of systems proposing ways to find the best or shortest path from point A to point B in different environments. Research, however, is lacking techniques to identify objects, especially in the context of a real-world search engine. The next chapter compares the four representations, used in the preliminary study, in terms of their usability in large scaled environments. Furthermore, the concept and design space of a real-world search engine is outlined in the next chapter.

3 Concept and Design Space

We have mentioned the term real-world search engine several times to this point but never provided a satisfying definition. This chapter provides a in depth analysis of the concept and design space of real-world search engines. As previously mentioned in the related work (see Chapter 2) this thesis is based on two previous projects. We use visual object recognition based on visual features as object identification mechanism as proposed in "Antonius" [Funk et al., 2013]. Furthermore, we use a subset of the location representations proposed in the preliminary study [Boldt et al., 2013].

3.1 Definition of a Real-World Search Engine

Search engines for the world wide web are well-established. Commercial systems like Google[1] or Bing[2] have evolved over the last years. Searching for information on the Web is as easy as entering a phrase as a search query to a website. The website is offering information about the query, roughly everything one could know about the sought phrase. Accordingly to search engines for the world wide web a search engine for the real-world is offering information about physical objects, e.g. their location. This is part of the Internet of things (IoT) [Ashton, 2011]. The IoT describes a system that identifies objects and their virtual representation, but in contrast to traditional data recording, humans do not need to manually record data. People have limited time, attention, and accuracy and therefore they are inferior compared to an automated recording of data. With the power of automated stocktaking people are enabled to track and count all available objects.

3.2 Basic Assumptions

As shown in the related work (see Chapter 2) existing systems that are able to identify objects, imply huge changes to the environment and therefore they are neither cost efficient nor scalable. Our main goal is to provide a system that works out of the box without major changes

[1] http://www.google.com/
[2] http://www.bing.com/

to the environment. We do not want to add any tags to objects or readers to a room or furniture.

To enable an out of the box behavior we identify objects by their visual features using camera streams. As we do not want to modify the environment, we do not add cameras to the environment but use wearable cameras. Keeping in mind that commercial head mounted displays (HMD), including a camera, (e.g. Google Glass[3]) are about to enter the market a system relying on wearable cameras becomes feasible. The camera images that need to be tagged with a geolocation will be analyzed by the system in real-time. If an object is detected, the current position of the object will be saved as well as the image.

To be able to add a geolocation to the image we need to determine the location of the camera. We showed several indoor-location technologies in the previously introduced related work (see chapter 2.2).

Unlike most of the proposed indoor location technologies we do not want to modify the environment. Therefore, we decide to use the infrastructure that had been already available and should be available at most places: WiFi. By measuring the signal strength of the available WiFi access points, adequate accuracy can be achieved [Woodman and Harle, 2008]. Therefore, we use WiFi indoor location technology to determine the position of the camera.

3.3 Selecting the Appropriate Location Representations

The focus of this thesis is to evaluate the representation of an object's location. Coordinates that are stored in a database need to be translated into a human-readable format. As shown in the related work (see Chapter 2) there was a preliminary study in which object representations were compared on room-level [Boldt et al., 2013]. Textual, 2D, and 3D representations were compared to an image-based representation in a lab study (see Figure 2.1). The main difference of this thesis to the preliminary study is that first the environment is a whole building and second the representations are shown on an HMD. The following chapter examines the four location representations depending on these two aspects.

The **textual** representation worked well on room-level, even thought subjective feedback of the participants discouraged the textual representation. With a growing area to be covered and more detailed zones the representation will become hard to understand. If a piece of furniture had been moved then the zones would have had to be changed accordingly. Therefore, we decided not to use the textual representation for an area greater than a single room.

The **3D** representation performed well on room-level and users were enthusiastic about it. However, creation of a detailed enough model of the environment is a highly work intensive

[3] http://www.google.com/glass/

job. The user needs to be provided a detailed model of the environment due to two reasons. One reason is that the relation of a visual location to a physical locations will be way more mentally demanding using a coarse 3D model compared to a detailed model including textures and a realistic size ratio. The other reason is that objects will be drawn on or inside pieces of furniture and therefore the furniture needs to be present in the model. Changes made to the environment like moving a piece of furniture need to be changed accordingly in the model. The coarse 3D model that was used in the preliminary study, was the main criticism of the 3D representation. In the preliminary study the 3D representation showed only coarse models of the furniture without any textures added to the model. Due to the work that is mandatory to realize a 3D model we did a brief research on how to create 3D models of a complete building and found several commercial providers of 3D laser scanners (e.g. faro[4]). Another possibility of 3D object scanning is the use of Kinect Fusion[5]. Both, Kinect Fusion as well as 3D laser scanners can create 3D models of the environment. However, we believe that creating and maintaining such 3D models will be highly uneconomic. Also, with a growing area that has to be covered the representation will become hard to understand. According to the preliminary study, users seem to like the 3D representation more than the textual representation. Nevertheless, we decided not to use the 3D representation in this project. We are discouraging everybody to use a 3D representation in the context of a real-world search engine due to the previously given reasons.

The **map** representation performed best in the preliminary study. In order to implement the map representation, a model of the environment is needed as well. Luckily, construction drawings of almost every building are available and should be sufficient. A basic design decision that need to be taken, regarding the map, is whether to shown furniture or not. If furniture is shown, the map needs to be adjusted with each environmental change. Without showing furniture the map is missing context but is more robust. In the preliminary study, the map was favored even though no statistic significance could be shown. Due to the easy implementation and robustness of the map we decided to use the map as a possible object location representation. The map is used without showing furniture.

The **last-seen image** representation requires no model of the environment and is independent of changes made to the environment. The hypothesis is that the last-seen image offers environmental context which enables a user to determine the position of a sought object. The hypothesis is supported by the findings of the preliminary study. Not only humans are able to perform localization based on visual features. Robots equipped with cameras can calculate their location by analyzing visual features of the environment [Se et al., 2002]. Human beings are more powerful in image recognition than a machine, especially in understanding the context given in a picture. Examples of human capabilities can be seen in the context of crowd sourcing (e.g. Amazon Mechanical Turk[6]). There are a lot of problems that are

[4]http://www.faro.com/
[5]http://msdn.microsoft.com/en-us/library/dn188670.aspx
[6]https://www.mturk.com

solved more efficient using human power than using artificial computation. An example for the power of crowd sourcing can be seen at “fold it”[7]. “Fold it” is a game everyone can play and simultaneously contribute to the scientific problem of protein folding. The last-seen image proved to be really powerful in the past. Therefore, we decided to use the last seen image for the outlined system.

Until now, an issue that has been concealed completely is the accuracy of the indoor positioning system. The textual, 3D, and map representation struggle with the accuracy of the indoor positioning system. For example: if an object lies in a shelf that consists of $5 * 5$ boxes, each of which are sized 50 cm in length and height. The object is in the middle box (3,3). WiFi indoor location has an accuracy of less than one meter (assuming the best case) and is only offering a 2.5 dimensional location (returns only longitude, latitude, and level of the floor but no exact height). The 3D representation cannot denote any position on or in the shelf because the accuracy is to coarse. In the best case the 3D representation could denote the shelf as being probably the furniture that contains the sought object. The textual representation and the map are not as sensitive for these kinds of errors. However, these representations tend to fail, if the location is too vague. In contrast, the last-seen image is robust against these kinds of errors.

In this section we made statements about the non functional requirements of each representation. We decided not to use the textual and 3D representation due to the unscalability, development effort, and lack of robustness. The map is used in the system because it was the most favored representation throughout the preliminary study. It offers scalability even though effort needs to be taken to scale the environment. Maps for the newly added environment have to be created, but the effort of creating maps from blueprints or even from scratch is manageable. Due to the scalability, low development effort, and robustness we decided to use the last-seen image system as well. The last-seen image offers an out of the box behavior that is fully model independent. The only downside of the last-seen image is that it did not perform best in the preliminary study, nevertheless it performed well.

3.4 User Stories

During the project we interviewed several people about the concept of a real-world search engine. We asked them about the possibilities and the use cases in which they would utilize the system. To provide a good understanding of the power of a real-world search engine these scenarios are shown in this section.

[7] https://fold.it/

3.4.1 Where Did I Put That?

As a user, I want to find a lost item so that I save time compared to a traditional search task.

This user story is the centric topic of this thesis. We spent most of our resources to provide satisfying solution to this user story. The concept of a real-world search engine is not limited to the type of the environment. It works at home, office, or even in warehouse environments. Especially in office environments, where a lot of parties are involved, important items can get lost easily. A real world search engine can tremendously support users in that context.

3.4.2 Where Did I Park?

As a user, I want to find my car so that I do not need to search the whole car park if I forgot where my parking lot is.

Finding a parked car can be solved using visual object recognition. Automatic license plate recognition is a widely researched topic [Chang et al., 2004]. Every car has a number plate that can be read by the system. The system can save the current position of the car. However, in general the use case for finding a parked car is outdoors. Outdoor environments are out of scope of this thesis. Nevertheless, extending the scope seems to make sense and should be researched with some caution in future work. The outdoor environment implies a lot of new questions about privacy and the usefulness of the last-seen image in outdoor environments. Especially in countries where people are sensitive about private data, the privacy issues of such systems will most likely raise a lot of issues.

3.4.3 Where Is the Next?

As a user, I want to find the next elevator so that I feel comfortable in unknown environments.

The elevator in this context can be substituted with almost everything e.g. conference room. When having an indoor-positioning system setup and running, it is not hard to setup points of interest. The Navigation could be done equally to the location representations of object search.

3.4.4 Where Is My Colleague?

As a user, I want to find my colleagues so that I do not need to search them manually.

With every user wearing a positioning device it would be easy to show the location of others to a user. As this would be highly privacy invasive a detailed permission system is needed.

However this idea is not new, sophisticated ideas for a permission system can be found for instance at “Find My Friends!”[8].

3.4.5 What Do I Need to Buy

As a user, I want to know what I need to buy so that I do not need to create a shopping list beforehand.

If all consumables are tracked, the system would be able to generate lists of any missing items. Hence, two possibilities arise, either create a shopping list or automatically order the missing items online. Extensions to this system are imaginable as well. The system could be coupled with a recipe database and create shopping lists according to the meals for the next days.

3.5 Design Space

The concept of a real-world search engine that tracks objects based on their visual appearance offers a lot of opportunities. It allows the system to be non-intrusive to the environment, even though it relies on the fact that everybody wears the system at least most of the time. No markers are needed and no furniture needs to be equipped with any kind of readers or sensors. However, the concept implies limitations as well. During the time we have been working on this project, we received a lot of questions especially regarding privacy. The following section addresses these concerns and limitations.

3.5.1 Privacy

The privacy concerns in the context of a real-world search engine that uses visual object recognition are twofold. One concern is that taking pictures of people and afterwards showing these pictures to other people is highly privacy invasive. The other concern is that people do not want others to retrieve information about their personal belongings or view pictures of their room taken in private. Therefore, guidelines need to be defined to protect the privacy of people as good as possible.

Showing the last-seen image of an object can be privacy invasive. As the last-seen image may contain some kind of sensible information. For example, a last-seen image may contain business contracts or a person who does not want to be recorded. Obviously, privacy can be restored at any time by taking off the system. However, in public rooms there may be

[8] https://www.apple.com/apps/find-my-friends/

several people wearing the system. Until everybody takes off the system, privacy cannot be guaranteed.

Automated disguise of people using algorithms is possible, e.g. “YouTube” proposed a technology that automatically conceals every face in a video by means of face blurring[9]. Event though these algorithms tend to fail if a face is not fully visible or in bad light conditions. Nevertheless, privacy can be improved using such technologies. However, in the context of a home or the work place, people will be able to identify their friends and colleagues without seeing their faces but because of other visual hints, e.g. their shoes, body language, or any other individual feature. Recent research [Rice et al., 2013] shows that the recognition of a person is not only possible due to face recognition but also on body features.

Additionally, we also propose an extension to the previously described face blurring mechanism. Nowadays, optical character recognition software is commonly used in a lot of applications like: passport recognition[10] or creating textual versions of printed books[11]. Automatic disguise of all characters recognized in the last-seen image increases the privacy even more.

People do not want others to watch pictures taken in their private rooms. Consequently, people do not want others to be able to track their private items. To address these issues, we propose to declare objects as well as rooms to different distinct privacy types. These types could be designed accordingly to the UNIX filesystem permissions[12]. We propose a hierarchical approach to declare an entity (in this context a room or object), to be accessible: globally, to a group, or only private. Global accessible entities can be accessed and viewed by everyone. Group accessible entities can only be accessed by users that belong to a certain group, like a department or a project. Private accessible entities can only be accessed by the owner of the entity. Accessible in terms of a room means that images taken inside the room are viewable. Accessible in terms of an object means that the object itself is searchable.

Some people have concerns regarding their privacy. They claim that uploading sensitive data to a third party does not leave a good feeling [Funk, 2012]. Therefore, we propose to make sure that the image processing and storage is only done using private servers. This ensures that the images are not accessed by third parties. Thus, we are able to reduce a lot of concerns people have regarding their privacy.

A system as previously described will always be privacy intrusive. However, we showed possibilities to further support a users’ privacy. Setting up a sophisticated permission system and sensible handling of data is highly recommend. These modifications can be done easily and provide a useful base for a real-world search engine. Automatically disguised images are not as easy to implement. Nevertheless, we highly recommend using disguise techniques in a

[9] http://youtube-global.blogspot.de/2012/07/face-blurring-when-footage-requires.html
[10] http://www.expervision.com/find-ocr-software-by-document-types/ocr-software-for-passport-processing-1
[11] http://www.gutenberg.org/
[12] https://www.freebsd.org/doc/handbook/permissions.html

real-world search engine. Yet, systems recording sensible data are only feasible if all concerned parties agree to the system.

3.5.2 Static vs. Mobile Cameras

Thinking about a system requiring camera images to provide any form of image processing one of the major questions is: “How to achieve an apropriate coverage of cameras?”. We have already explained the concept of using wearable HMDs with included cameras to retrieve images. The following chapter motivates why statically mounted cameras are inferior to mobile cameras.

Mounting cameras to the environment to be able to track the available items implies mounting several cameras in each room. To be able to film inside a drawer and recognize objects, cameras need to be mounted above drawers as described in Drawerfinder [Komatsuzaki et al., 2011b]. Static mounted cameras always have a blind angle. No objects will be detectable in the blind angle. The only possibility to get rid of blind angles is mounting even more cameras. When mounting cameras to the environment, users may feel observed [Funk, 2012].

Using a mobile camera does not imply blind angles. Less cameras will be used, as each camera is moved throughout the environment. One camera per person will be enough. As objects are only moved by humans almost every movement will be tracked by the camera [Pederson, 2001]. Obviously, the concept is not working if somebody is dropping an object without looking at it and therefore filming it with the camera. The next person passing the object records it again. In addition, mobile cameras are able to film places that are complicated to film with static cameras, e.g. recording the contents of a drawer.

It can therefore happen that objects are registered at the wrong place for a limited time frame. Faults are recovered without the need of human intervention. Thus, the proposed system is robust and provides eventual consistency [Vogels, 2009].

We believe that a last-seen image recorded in point of view is easier to understand compared to a last-seen image recorded from the ceiling. We do not want to modify the environment or that users feel observed. People who care about privacy may take off the system at any time to restore privacy. For this reasons, using mobile cameras is the better choice for the proposed concept.

3.5.3 Object Recognition

The most valuable asset of using object recognition is the waiver of using tags attached to objects. Objects can be added to the environment by taking a photo. It does not even need to be an actual photo of the object. Most of the time an image found on the web is sufficient. It

may even be possible to setup a community driven reference database containing information about most daily items including reference images to identify these objects.

Visual object recognition implies limitations that need to be addressed in order to use a real-world search engine based on visual object recognition in a productive environment. It is possible that objects are not recognized or even worse that they are mistakenly recognized. Recognizing objects on their visual features as well as their appearance does not show as accurate results as tag based approaches like RFID. Visual object recognition is a growing field and we believe that a lot of advantages will be made to existing algorithms. Results proof that visual object recognition is adequate for the use with a real-world search engine in a lab study [Funk et al., 2013].

Nevertheless, object recognition shows a major flaw that needs to be overcome. There are objects that look alike or are the same, e.g. laptops or smartphones. There is no perfect solution to differ between these objects without manipulating them in their visual appearance, like adding visual markers. Techniques like mapping objects to people and analyze either who filmed or who interacts with an object are prone to error.

Object recognition is a highly promising technology to setup a real-world search engine. We believe that, until almost every object comes naturally equipped with a RFID tag or something similar, visual object recognition is the only feasible technology to setup a real-world search engine in a large scaled environment.

3.5.4 Centralized vs. Mobile Object Recognition

Visual object recognition is a performance intensive task. Our long-term goal is to provide a scalable productive real-world search engine working out of the box. However, the decision whether to use a centralized or a mobile object recognition cannot be answered without sufficient experiments and tests. Using an HMD for object recognition retains our goal of an out of the box behavior. In contrast, a centralized service is able to perform more complex computations and therefore better object recognition.

Mobile phones are suitable for visual object recognition in real-time [Henze et al., 2009]. The computational performance of the mobile phone that has been used by Henze et al. is inferior to the computational performance shown by current HMDs. But there are also possible downsides of running the object recognition on mobile devices. With the growing number of reference images the performance limits will be reached at some point. Consequently, mobile object recognition is not scalable. Furthermore, using object recognition on mobile devices highly utilizes the mobile devices. Therefore, the battery is drained. Current HMDs suffer on

battery life. According to current reviews, the battery of Google Glass is wiped after one hour of video recording[13].

Using a centralized architecture performing visual object recognition may break with the out of the box behavior. There are two scenarios available. Using a public image processing service (e.g. recognize.im[14]) does not break with the out of the box behavior. But there are users who do not want their pictures to be uploaded to a third party service due to privacy reasons (for more information see Section 3.5.1). Setting up a centralized service in the local environment implies extra work that needs to be done before being able to use the system. A centralized service also implies high costs due to the environment that needs to be either setup or paid by usage. A centralized service can nevertheless offer a highly scalable environment and complex computations to evaluate the camera streams.

Using mobile object recognition breaks with the scalability of the system. Using a centralized service breaks with the out of the box behavior of the system. A hybrid system could solve the issue. Using a mobile device for object recognition seems to be suitable for small scenarios. In large scaled environments the usage of a centralized service seems to be convenient.

3.5.5 Last-Seen Image

The concept of a system relying on visual object recognition implies not only the existence of images taken of the environment, but also of sought objects. Therefore, the term last-seen image was created. The last-seen image shows the most recently captured image of an object including the environmental context of the surroundings.

We conducted two previous studies to proof the usability of the last-seen-image. One study compared the last-seen image to 3D, map, and textual representations on room-level [Boldt et al., 2013]. The last-seen-image was shown to be equally usable compared to the other representations. As a followup the authors conducted an informative user study to proof the useful amount of context, the last-seen-image is offering. Several objects were photographed in an office environment. These photographs where then presented to participants that are familiar with the environment. The office environment consists of 12 rooms and a hallway. People could relate 93% of the pictures to the correct location. One image, only showing a shelf and its content, was related wrong by more than half of the participants. An explanation could be the missing environmental context in the picture.

Office environments differ from home environments. In office environments there is a lot of look alike furniture and rooms. In contrast, in home environments there are a lot of different looking rooms. The previously described issue, of a wrong related last-seen image, can be tackled. We propose three modifications that can improve the last-seen image. The last-seen

[13] http://www.techradar.com/reviews/gadgets/google-glass-1152283/review/7
[14] https://www.recognize.im/

Figure 3.1: The difference of a last-seen image providing almost no contextual information compared to a last-seen image combination taken from three distinct angles providing a lot of context. (a) A last-seen image providing almost no context. (b) Three last-seen images creating a surrounding view of an object.

image can be extended by some sort of compass alike indicator that shows the direction and distance towards the sought object. The second and third improvement only works if there is more than one last-seen image (of the same object at the same place) available. The second improvement is to select the best last-seen image. The third improvement comprises a merge of several last-seen images.

It is hard to rate the quality of one last-seen image compared to another last-seen image. A lot of items remain at the same position for some time, since nobody is moving them constantly. Therefore, there will be several last-seen images of the same item at the same position. Picking the latest image is not always the best choice.

Some HMDs feature a compass. When taking a photo the current orientation of the camera can be stored for each image that was taken. If there is more than one last-seen image of the same object at the same position available, a representation of several last-seen images can highly increase the usability of the last-seen image (see Figure 3.1). Three last-seen images would create some sort of a surrounding view of the object. The chance that at least one image contains vital contextual information to determine the position of the object is increased.

We conducted a short interview to find out more about the last-seen image. We wanted to find out how users rate different last-seen images against each other. The other information of interest was whether users would want one high quality last-seen image or a collection of several last-seen images. We therefore showed the participants different last-seen images of the same object at the same position and asked them which image was their favorite one. Afterwards, the participants were asked whether they prefer their favorite image compared to a collection of three last-seen images taken from different angles. All images were taken in a home environment.

We recruited 12 participants (7 female, 5 male), who agreed to take part in our interview. Their age was between 21 and 48 years ($M = 35.42, SD = 10.26$). None of the participants received a financial compensation. The participants had a wide range of professions. There

were students, office workers, a hotel manager, a musician, and a project leader. All of the participants know the environment in which the images were taken.

The interview showed that most of the participants favored a last-seen image that was taken from about 1 to 2 meters distance of the sought object. Last-seen images showing the sought object in the center of the picture were favored. In contrast, last-seen images showing the sought object at the edge of the image, but offering a lot of environmental context, were disliked in favor of the previously described image. Participants who were familiar with the environment preferred last-seen images taken closer to the sought object since the given context was still sufficient to identify the location of the sought object. If a valuable last-seen image, providing enough contextual information, was available, the participants disliked a merged representation showing the sought object from different angles. Participants stated that the surrounding view would be too much information.

The last-seen image is a very powerful technique of representing an object location. It can be improved by an algorithm that automatically selects the best last-seen image. If the user is unable to identify the location of the sought object by using a single last-seen image, it is recommend to use a collection of several last-seen images that shows the sought object from different angles.

In this chapter we chose the last-seen image and the map representation to be convenient for the use in a whole building. We showed scenarios that can be solved by a real-world search engine. Additionally, we provided advice about the implementations of these scenarios. The last part of this chapter deals with the design space of a real-world search engine based on visual object recognition that uses the last-seen image as favorite representation.

We propose a real-world search engine based on visual object recognition. The HMD provides a map and last-seen-image representation to guide a user towards a sought object. A prototypical system enabling users to search for objects, using the two representations, is described in the following chapter.

4 Prototype

This chapter describes the prototypical implementation of the previously described system (see Chapter 3). The system provides a map and the last-seen-image representation on a head-mounted display (HMD). Both representations have been introduced (see Chapter 2) and discussed (see Chapter 3). We do not want to include any visual object recognition due to two reasons. One reason is that visual object recognition has been explored in previous research [Funk, 2012]. The other reason is that this project focuses on the location representations and not on object recognition. The field of representing an objects' location is lacking research.

4.1 Hardware

The perfect HMD for the system would be Google Glass[1]. Unfortunately we are not in possession of a developer version of Google Glass (at the time, writing this thesis, Google Glass has not yet been released). Therefore, we had to find an adequate replacement. We use an Epson Moverio BT-100[2] (hereafter referred to as "Moverio") as a replacement for Google Glass. It offers a head-mounted see-trough display. The Moverio lacks a bunch of features and sensors that are required in order to accomplish the task of navigating users towards objects. As mentioned before, we want to rely on WiFi indoor positioning, therefore a good WiFi scan and refresh rate is mandatory. According to our tests, the Moverio does not offer either. The Moverio does not provide any acceleration sensors, gyroscope, or compass[3]. We decided not to use the Moverio as the positioning device but use it as the user interface device. We use a Samsung Nexus S[4] (hereafter referred to as "Nexus") to retrieve the indoor position. It supports everything that is necessary: acceleration sensors, gyroscope, compass, and a good WiFi scan rate. The Nexus streams the current location to the Moverio using the User Datagram Protocol (UDP). The whole system can be seen in Figure 4.1.

[1] http://www.google.com/glass/
[2] https://www.epson.com/cgi-bin/Store/jsp/Moverio/Home.do
[3] Epson announced a successor to the Moverio BT-100, the BT-200 that includes a front facing camera, Bluetooth, a gyroscope, and several other features. See: http://www.epson.com/moverio
[4] http://www.samsung.com/us/support/owners/product/GT-I9020FSTTMB

Figure 4.1: The prototype uses an Epson Moverio BT-100 as user interface device and a Samsung Nexus S to retrieve the position using WiFi indoor positioning.

4.2 WiFi Indoor Positioning

Due to the fact that there are a lot of commercial providers available that offer indoor positioning we decided not to implement our own indoor positioning application programming interface (API). This decision was taken because this thesis focuses on object location representations and not on indoor positioning. Indoor positioning is mandatory to support navigation towards a sought object, as well save the location of an identified object. We created a comparison of several WiFi indoor positioning providers to select an appropriate one (see Table 4.1).

We sent out several requests to non-free indoor positioning providers, asking for permission to use their system for research purpose. Navizon[5] generously allowed us to use Navizon Indoors free of charge. This allows us to do a comparison between a non-free and a free provider to decide which one is preferred in our environment. The environment for the indoor positioning

[5]http://navizon.com/

Name	Accuracy	Notes	URL
Qubulus	<1m	Not free, requires personal contact	http://www.qubulus.com/
indoo.rs	>2m	Free with ads, well documented, measurement is done on Windows/Unix	http://indoo.rs/
Navizon	<1m	Not free, Android and IOS	http://navizon.com/
Redpin	room-level	Opensource, Android and IOS	http://redpin.org/
WiFiSLAM	2.5m	Acquired by Apple, March 2013	Website has been taken down
Google	<5m	Free, indoor location implemented in the Android App, more than 10.000 buildings mapped	https://maps.google.com/help/maps/indoormaps/
Wifarer	room-level	Not free, in app positioning	http://www.wifarer.com/
Polestar	2-5m	Not free, Combination of WiFi and GPS	http://www.polestar.eu/
QBengo		Not free, not much information available	http://www.qbengo.com/
IndoorAtlas	3m	Free, measures magnetic variations of the building	https://www.indooratlas.com/

Table 4.1: Comparision of indoor positioning providers: these data was last checked May 1st, 2014

is the hciLab[6], an almost 500 square meters ground area consisting of three floors. The whole building consists of a lot of concrete and due to the information technology background there are a lot of interference waves, created by a huge mass of electronics. The indoor positioning system thus needs to be robust to withstand these challenges.

We set up a small test scenario in a hallway that is about 20 meters in length and two rooms that adjoin the hallway. We decided to compare indoo.rs[7] to Navizon. The reason why we chose these two competitors is the brilliant documentation that made setting them up easy. Other free indoor positioning providers (see Table 4.1) are poorly documented so that even setting them up for our test scenario would be hard work.

While calibrating the two systems, we discovered the first differences. Indoo.rs requires calibration via a laptop. The calibration is done by standing at one point for about half a minute measuring the signal strength of the available WiFi access points. As there are several fingerprints needed to allow exact indoor positioning, the calibration of indoo.rs took roughly

[6]http://www.hcilab.org
[7]http://indoo.rs/

an hour for the test scenario. Navizon's calibration is done using a smartphone to collect fingerprints while walking. Routes can be set up on a map and followed in the physical world to measure the WiFi access points along this route. The calibration of Navizon took us less than half an hour.

We compared the accuracy and the update interval of Navizon to indoo.rs. The accuracy as well as the update interval of indoo.rs were inferior to the results provided by Navizon. Therefore, we decided to use Navizon for providing an indoor location for our prototype. These tests took place in November 2013 at the hciLab and therefore represent the development state of these two indoor positioning providers at that time.

4.2.1 Integration of Navizon

As mentioned before the Moverio shows a poor WiFi scan rate and therefore we were in great need of a device that offers a proper WiFi scan rate to be able to retrieve the current position. We did a quick and informal comparison of a Samsung Nexus S and a Samsung Galaxy S3 using the Navizon App[8] and found that the Nexus offers a better WiFi scan rate compared to the Galaxy S3. Thus, we use the Nexus as positioning device.

The whole hciLab was calibrated using the Naivzon App. The integration of Navizon indoor positioning service to our existing Android App was straight forward. Navizon offers a representational state transfer (REST) interface to retrieve the current position without requiring to implement a software development kit (SDK). We had to measure the signal strength of the available WiFi access points and send them to Navizon. Navizon replies with the geodetic coordinates (e.g. $48.746889N, 9.107988E$) of the current position. As the positioning part of the system is done using the Nexus, the Nexus streams the received coordinates including the current compass data to the Moverio. The evaluation of the coordinates is done by the representations on the Moverio.

4.3 Last-Seen-Image Representation

The last-seen-image representation shows a picture of an object in the context it has been recognized by the system lastly. Surrounding furniture and other objects are shown to support the localization of the sought object. Before showing an image we want to navigate the user to room-level. This was proposed in the design space (see Chapter 3.5.5), due to possible ambiguous context shown in the last-seen image. Indicating a user to switch floors is done by showing a stairways symbol including an up or down arrow (see Figure 4.3). If the user is on

[8] https://play.google.com/store/apps/details?id=com.mexens.android.navizon

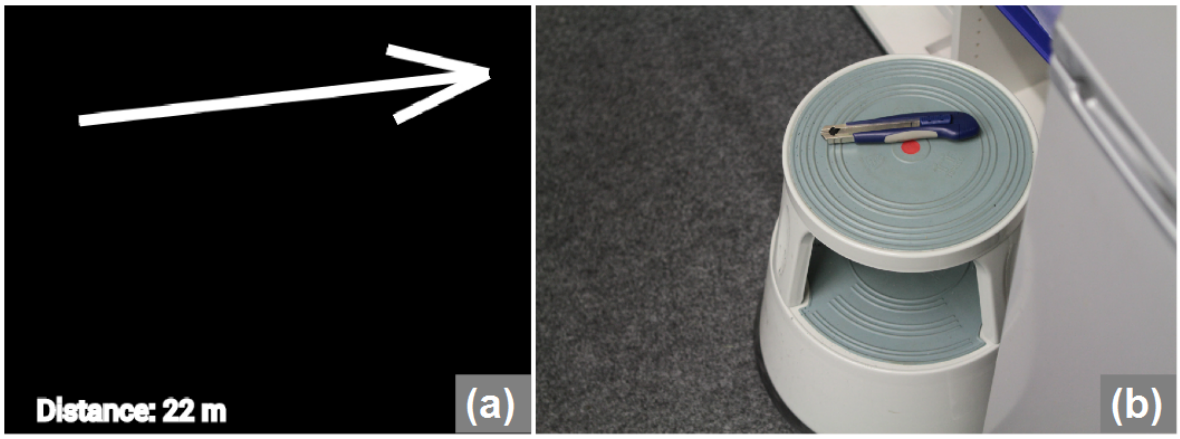

Figure 4.2: Two screenshots of the system using the last-seen image representation. (a) An arrow pointing towards the sought object and showing a distance to the sought object. (b) A last-seen image representation showing the position of a cutter including environmental information.

the correct floor but not in the correct room an arrow pointing towards the object including a distance indicator is shown (see Figure 4.2).

The system is calculating the distance as well as the angle (referred to as bearing) of the arrow by using the geodetic coordinates of the current position as well as the geodetic coordinates of the sought object. To be able to calculate the bearing of the arrow the current azimuth is needed as well. The azimuth is retrieved using the build in compass of the Nexus. Unfortunately it showed that the compass is vulnerable to failures in indoor environments. Since we have not implemented the object recognition part of the system, the coordinates of the sought objects are at a fixed position. At the current state the focus is on comparing the representations to each other.

If a user comes closer than 5 meters towards the sought object the last-seen-image is shown (see Figure 4.2). Showing the last-seen-image in a 5 meter radius proved to be a good solution during our tests of the system. Due to accuracy reasons and to prevent distraction the arrow is hidden while the last-seen-image is shown. The indoor-location system is not sub-meter accurate and therefore it can happen that the arrow is pointing into a wrong direction.

The sought object management including the location as well as the last-seen-image is loosely coupled to the main program. As we proceed it is a good idea to completely decouple the sought object management from the main program to enable manipulation of the data by a third party. If somebody else has, for instance, moved an object you are looking for, you would want the location to be updated, so the object is still findable. Therefore, it is a good idea to move the sought object management to a webserver. But because of the fact that we focus on the location representation we left this point open to future work.

4.3.1 Calculating the Distance towards the Sought Object

To represent and calculate data of the real-world we need a model of the world. When calculating distances of two points using coordinates we need a model representing the surface of the earth. This is also known as calculating the geographical distance. While the earth is round, it is neither a perfect sphere nor a perfect ellipsoidal. There are three common abstractions: flat surface, spherical surface, and ellipsoidal surface. Spherical and ellipsoidal abstractions of the surface are more exact than using the flat surface abstraction.

Nevertheless, we decided to use the flat surface abstraction due to several reasons. In general the flat surface abstraction starts to show noticeable errors for distances larger 12 miles (20 kilometers) [Sinnott, 1984]. The biggest distance we want to calculate is about 100 meters. For small distances the use of the flat surface abstraction is recommend, because it can be computed very fast.

Using the flat surface approach, we ended up using basic Euclidean geometry. The shortest path of two points with known coordinates in a two-dimensional coordinate system is a straight line that can be calculated using the Pythagorean theorem.

In the following formulas we use the here declared variables. All coordinates are in radians. θ is the symbol used for latitude values. $\Delta\theta$ is the difference of the two latitude values of which the distance shall be calculated. λ is the symbol used for longitude values. Accordingly, $\Delta\lambda$ is difference of the two longitude values of which the distance shall be calculated. θ_m is the arithmetic mean of the two latitude values. R is the radius of the earth in meters (6,371,009 meters). D is the difference of the two coordinates in meters.

Using regular geodesic coordinates, one needs to be aware of the fact that longitude and latitude vary in their scale [Snyder, 1997]. This is known as equirectangular projection and defines the scale between the latitude and the longitude to be:

(4.1) $a = \Delta\theta \qquad\qquad b = \cos(\theta_m)\Delta\lambda$

The Pythagorean equation is defined like this:

(4.2) $c = \sqrt{(a)^2 + (b)^2}$

Now substituting the a and b:

(4.3) $c = \sqrt{(\Delta\theta)^2 + (\cos(\theta_m)\Delta\lambda)^2}$

The c-value is now a value calculated by two coordinates in radians, to convert that cryptic value to a human readable value we need to multiply the value by the radius of the earth. And concluding the final formula:

(4.4) $D = R\sqrt{(\Delta\theta)^2 + (\cos(\theta_m)\Delta\lambda)^2}$

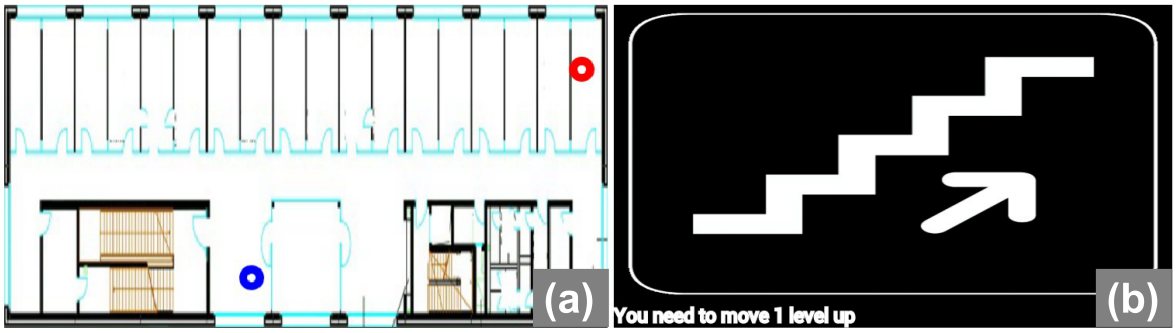

Figure 4.3: Two screenshots of the system using the map representation. (a) The map representation with the sought object denoted as red dot and the current position denoted as a blue dot. (b) A staircase symbol indicating the user to move one level up.

4.3.2 Calculating the Bearing to the Sought Object

A Bearing is defined as the angle between our forward direction and the direction from our current position to another object (sometimes referred to as forward azimuth). So the bearing is the angle of the arrow that points towards the sought object. The bearing can be calculated by using the Haversine formula [Veness, 2012] as follows:

$$B = \arctan\left(\frac{\sin(\Delta\lambda) * \cos(\theta_2)}{[\cos(\theta_1) * \sin(\theta_2)] - [\sin(\theta_1) * \cos(\theta_2) * \cos(\Delta\lambda)]}\right) \tag{4.5}$$

The variables in the Haversine formula are defined as follows. All coordinates are used in radians. θ is the symbol used for latitude values. θ_1 is the latitude value of the current position. θ_2 is the latitude value of the sought objects. λ is the symbol used for longitude values. $\Delta\lambda$ is the difference of the two longitude values of the sought object and the current position. B is the bearing towards the sought object.

4.4 Map Representation

The map representation can be seen as the traditional location representation for indoor as well as outdoor location representation in general. The basic idea in most implementations is to denote the current position and the destination on the map. We created maps for each floor by using the blue-prints of the hciLab. A map rotation was omitted intentionally, due to two reasons. One reason is that we base the representations on the preliminary study that does not include map rotation, since the representation was given on a desktop PC. The other reason is that map rotation is vulnerable to compass errors which happen in indoor environments. The current position is denoted as a blue dot on the map and the destination is denoted as a red

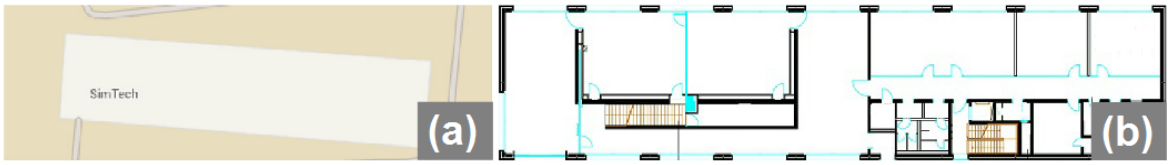

Figure 4.4: Comparison of the alignment of the hciLab and the map representation given in our app. (a) A screenshot of the HCILab taken of Google Maps aligned north up. (b) The map of the ground floor as used in our app to map coordinates. Both screenshots are rotated in different angles therefore the mapping of coordinates to the map needs to be adjusted accordingly.

dot (see Figure 4.3). If the user is on a different floor a stairwell icon is shown that indicates the change of the floor.

4.4.1 Mapping Coordinates to the Map

Mapping coordinates to the map means that for every given coordinate a position on (or outside) the map can be determined. The map representation's base function consist of showing coordinates on a map. Our initial approach was to take a reference point on the map with known coordinates. Determine how many pixels the given coordinates distinct from the reference point. Map the coordinates accordingly to the map. Unfortunately, it is not that easy.

The map of Google Maps is aligned north up (see (a) in Figure 4.4). In a north up aligned map coordinates can be mapped as previously explained. If a map is not aligned north up it is rotated (see (b) in Figure 4.4). Mapping coordinates as previously described is not correct if the rotation is ignored.

In the following discussion we use notions as used in a two-dimensional Cartesian coordinate system. The x-value (abscissa) is the horizontal value and the y-value (ordinate) is the vertical value. The latitude is the synonym for the x-value. Consequently, the longitude is the synonym for the y-value. To further ease the understanding, we use left as synonym for west and top as synonym for north.

We propose to use two different reference points. One reference point is on the left edge. The other reference point is on the top edge. Both can be calculated following the same schema. Therefore, we present only the calculation for the top reference point.

The top edge of the map (the mapping is done in (b), see Figure 4.4) is rotated. Thus, we need to calculate the y-value for a given x-value. The y-value of the top changes in some grade.

Therefore, we can define the top edge as a first degree polynomial. The polynomial shall return the y-value ($f(x)$) for a given x-value. Remember a first degree polynomial looks like this:

(4.6) $f(x) = ax + b$

b is the y-value of the top left corner. x is the difference of the x-values of the top left corner of the building and the x-value of the current position. a is the grade of the y-value variation between the top left corner and the top right corner.

Substituting these values using the previously used symbols for coordinates, we conclude with the following formula:

(4.7) $f(\lambda) = \frac{\theta_{tr} - \theta_{tl}}{\lambda_{tr} - \lambda_{tl}} * (\lambda - \lambda_{tl}) + \theta_{tl}$

λ is the longitude value of the position to be mapped. tl is the top left corner of the building. tr is the top right corner of the building. λ_x is the longitude value of the x corner of the building. θ_x is the latitude value of the x corner of the building.

Now that we know the top reference point as well as the left reference point depending on the current coordinates, we can map the coordinates to the map. This is done by calculating the difference of each pair of coordinates. We are left with calculating how many seconds[9] equal one pixel on the map. Thereafter, we can easily calculate the position of the given coordinates on the map. This technique again uses the flat-surface model as described previously (see Section 4.3.1).

[9]A second is a unit for measuring coordinates. For more information on representing coordinates see ISO 6709.

5 Evaluation of the Last-Seen Image

As described in the related work (see Chapter 2.3), there was a preliminary study that compared the last-seen image, textual, map, and 3D representations on room-level. We decided to use the map and last-seen image representation (see Chapter 3). These representations have been implemented for the use on a head mounted display (HMD) as described previously (see Chapter 4). In the following chapter the map representation is compared to the last-seen image representation in an experimental evaluation. The representations are compared in a building consisting of three floors.

The study described hereafter was conducted using the wearable prototype, as described in the previous chapter (see Chapter 4), consisting of the Epson Moverio BT-100 and the Nexus S. The Moverio is an HMD that shows the two representations to the user. The participant was followed by the experimenter who used the Nexus to stream the location to the Moverio. Overall, the study consisted of six search tasks, three of each using one representation.

The results of the study indicate the use of a combination of the last-seen image and the map. The map is more suited to find the way towards an object. On room-level the last-seen image showed to be more suitable. Overall, hard hidden objects were found faster using the last-seen image compared to the map approach.

Hypotheses
Our hypotheses is that using the last-seen image representation in a whole building is faster and less mentally demanding than using the map. This results arise due to the context the user is given with the last-seen image. People who are familiar with the environment can figure out the location of an object by viewing the last-seen image [Boldt et al., 2013]. A sophisticated discussion about the last-seen image can be found in the concept (see Chapter 3.5.5).
(**H1**) Using the last-seen image is more efficient than using the map representation.

5.1 Method

The study was conducted in a building consisting of three floors. Overall there were eight objects to search for. Two objects were dedicated to an introductive search for each representation. Figure 5.1 shows the six objects that were dedicated to actual searching.

Figure 5.1: The sought objects used in the study. Each object was searched for one after another. A contextless picture of the currently sought object was showed to the participant before each search task. The names of the objects from left to right are: cutter, remote control, water sprayer, phone box, vacuum, and first aid kit.

5.1.1 Design

This study used a repeated measures design [Vonesh and Chinchilli, 1996]. There were two independent variables. One was the representation with two levels: the last-seen image and the map. The other was the sought object with six levels. Each of the sought object is a level. A condition consisted of three search tasks. Thus the experiment comprised of six unique search tasks. The conditions were performed in a counterbalanced way using Balanced Latin Square [Bailey, 2008]. The time to find an object was measured for each search task. It is referred to as task completion time. It was measured from the moment on, in which a participant started to movie until the sought object was touched. After finishing three search tasks of one condition, the participant was asked to fill out the NASA Task Load Index (TLX) [Hart and Stavenland, 1988] and the System Usability Scale (SUS) [Brooke, 1996] questionnaires. We eliminated the pairwise comparisons of the Nasa TLX. This is often referred to as Raw TLX [Hart, 2006]. There were three dependent variables: task completion time, SUS, and TLX. Additional qualitative feedback was collected at the end of the user study.

5.1.2 Participants

We recruited 16 study volunteers (9 male, 7 female), who agreed to take part in our study. Their age was between 15 and 42 years ($M = 25.31 years, SD = 6.55 years$). Most of the participants were students, studying computer science. There was one pupil and an administration secretary. None of the participants received a financial compensation. At the beginning of the study the participants were asked to state how often they are searching for an object and are not able to retrieve it immediately. The results show that 6 of the participants stated to search on a daily base, 14 participants stated to search at least once per week.

5.1.3 Apparatus

The study was conducted using the wearable prototype described beforehand (see Chapter 4). The location was manipulated in a wizard-of-oz manner [Dahlback et al., 1993]. This decision was taken, because the WiFi location appears to be unstable from time to time. So it can happen that the location jumped more than 10 meters or was several meters off. Even though most of the time the position's accuracy was better than 3 meters, we did not want the user to be influenced by an inaccurate position. Further, wizard-of-oz enables comparable conditions for all participants. So we introduce a wizard-of-oz application to manipulate the location, level and azimuth. The wizard-of-oz application consisted of a map to set the location, radio buttons to set the level and a rotary knob to set the azimuth (see Figure 5.2). The task completion time was measured using a stopwatch. The contextless pictures that were shown to the participant in the beginning of the search task. The pictures were printed A4 pictures, taken in front of a white background. Each printed picture contains only one object. All forms and questionnaires that have been used can be found in the appendix (starting at page 65). Throughout the study we used the following forms: a consent form, demographics, SUS, TLX, and final questions.

5.1.4 Procedure

After the participants were welcomed and introduced to the main topic of the experiment, they were asked to fill out the consent form. They were also asked to fill out basic demographics and introductive questions. The introductive questions covered the search habits of the participants. Afterwards, the experimenter introduced the participant to the wearable prototype. Special care needed to be taken if the participant was wearing glasses, since glasses tend to disturb the view trough the HMD.

To become familiar with each representation, the participant was asked to search for one object using our prototype. The preliminary search was not measured and the experimenter guided the participant through the search. When the participant felt comfortable using the representation the actual search task was started. Every search task followed the same pattern

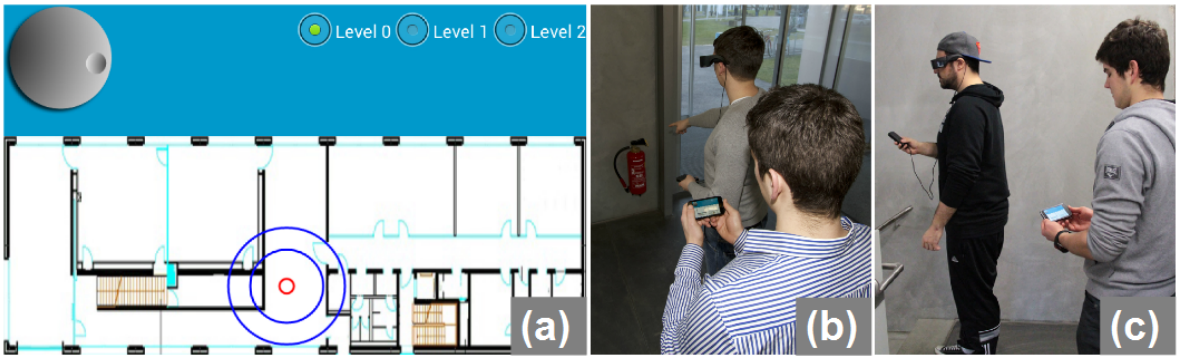

Figure 5.2: (a) A screenshot of the wizard-of-oz application. There is a rotary knob to set the azimuth as well as buttons to set the current level. The most important part is the map of the current level. Using the map the experimenter is enabled to set the current position of the participant. (b) A participant is about to find an object by touching it. The task completion time was measured until the object is touched. (c) The participant is wearing the prototype to search for an object. The experimenter is using the wizard-of-oz application to manipulate the location, azimuth, and level.

and started at the exact same position, on the second floor in the middle of the main hallway. Therefore, every task completion time measured for one object is comparable. Since the distance from the start to the same object is always equal. After naming the sought object, the participant was shown a contextless picture of the sought object. Thus, every participant knew exactly what the sought object looks like. In follow of the introductive section, the participant was shown the representation. As soon as the participant started to move away from the starting point, the experimenter started the time measurement. Due to the fact that this study was conducted using a wizard-of-oz design, the experimenter accompanied the participant in about one meter distance to be able to adjust the current position as accurate as possible. When the sought object was touched by the participant the task completion time measurement was stopped. After each condition was completed the participant returned to the starting point.

After the participants found three object using one representation they were asked to fill out the SUS and the TLX questionnaires. At the end of the study, participants were asked for additional qualitative feedback.

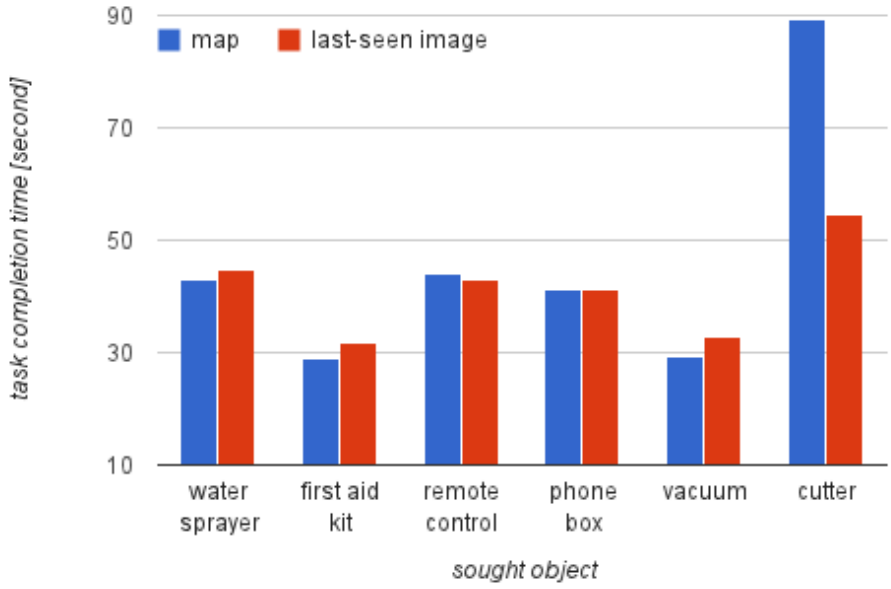

Figure 5.3: The diagram shows the task completion time of each object, measured in seconds. The task completion time of the cutter is significantly lower using the last-seen image compared to the map representation.

5.2 Results

A participant needed about 45 minutes to participate in our study. We experienced a huge variance in the completion time of the study. Some participants finished after less than half an hour, others needed more than an hour. The main reasons for the variance in the task completion time was that some participant rushed trough the study and gave almost no qualitative feedback. In contrast, other participants were laid back and answered the qualitative feedback in detail.

We compare three scores: the SUS score, the TLX score and the task completion time. In the following section we report and analyze these measures in detail. The abbreviations follow the recommendations of the APA Publication Manual [Association, 2012]. We start by reporting the task completion time.

We decided to only compare the task completion time for each sought object and not for each condition, since the objects are hidden in different distances and difficulties across the building. The results show that the task completion time of finding an object using the map ($M = 41.37, SE = 1.78, SD = 11.97$) is slightly increased compared to using the last-seen image ($M = 45.96, SE = 3.53, SD = 24.46$). A paired t test on the task completion time of each object could not reveal a significance for five objects. But using the last-seen

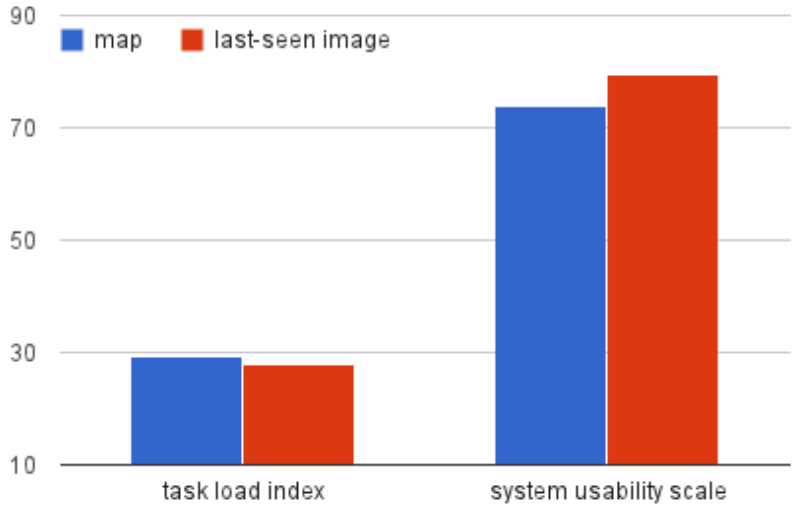

Figure 5.4: The diagram shows the results of the task load index (TLX) and the system usability scale (SUS) for each representation. We could not reveal a significant difference using either one of the representations.

image representation was significantly faster than the map representation for the cutter $(t(7) = -3.64, p < 0.01, r = -0.67)$. The cutter has the highest mean $(M = 71.91, SD = 25, 5)$ compared with the others (see Figure 5.3). The mean task completion time for finding the cutter using the map $(M = 89.13, SE = 8.24, SD = 23.31)$ is significantly higher than using the last-seen image $(M = 54.69, SE = 4.64, SD = 3.13)$.

The results of the SUS indicate that the last-seen image $(M = 79.38, SE = 1.94, SD = 13.44)$ was easier to use compared to the map $(M = 73.9, SE = 2.41, SD = 16.69)$ (see Figure 5.4). We could not show a significant difference for using either representation, by using a paired t test to analyze the given results $(t(15) = 1.59, p = 0.13, r = 0.17)$.

Comparing the task load using Nasa TLX the last-seen image $(M = 28.06, SE = 5.02, SD = 20.09)$ was less demanding than the map $(M = 29.25, SE = 4.22, SD = 16.87)$ (see Figure 5.4). Again a paired t test could not show a significant difference using one of the representations $(t(15) = -0.25, p = 0.81, r = -0.03)$.

One of the questions we asked the participants is: “what is your favorite representation”. The results show that 9 participants stated that the map was their favorite, 5 voted for the last-seen image and 2 thought both were equal.

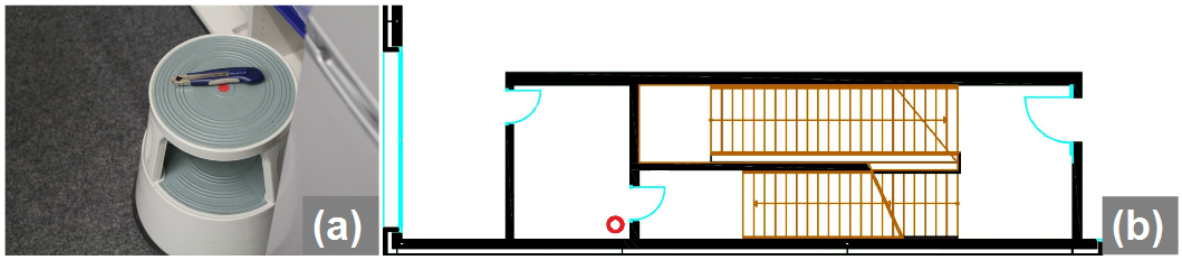

Figure 5.5: The difference between the cutter's location represented as last-seen image (a) and map (b). (a) The cutter's last-seen image shows the cutter to be close to the ground. A person who knows the environment can also detect that the cutter is in one of the storerooms. (b) The cutter is in a room behind the stairs. Some participants were irritated and started to climb the stairs.

5.3 Discussion

The results of this study are not promoting either one of the representations to be superior to the other. The mean of the SUS and the Nasa TLX score of the last-seen image are slightly better compared to the map, but not significantly. The mean of the task completion time also promotes the last-seen image, however, significant differences could only be found for one object. In contrast, the map was promoted the favorite representation by the participants. The findings do not fully support our hypotheses. Especially for searching trivial hidden objects, no significant difference can be shown. Yet, searching for harder hidden objects, the last-seen image is faster than using the map and therefore supports our hypotheses. Most of the objects are trivially hidden (lying on a couch or a table) and therefore can be located without the need of the additional context provided by the last-seen image. Objects that are not in the field of sight benefit from the context.

The only object that showed to be significantly faster using the last-seen image compared to the map representation was the cutter (see Figure 5.3). The cutter lay in a room behind the stairway and was positioned close to the ground (see Figure 5.5). Three participants were irritated and started to climb the stairs. They searched in the area of the stairs, even though the object lay in a completely different room. The context presented in the last-seen image would most probably prevented the participants to search in the wrong area, as the room looks like a storeroom and not like a stairway. Almost all participants that used the map representation had problems to identify the sought object on room-level. They moved even further, into the next room, because they were not able to identify the object. None of the participants that used the last-seen image moved to the next or wrong room when searching for the cutter.

The subjective feedback points out that people like to plan their route towards the object using the map. Participants are enabled to see their progress and can easier orientate themselves

in the building. The participants missed the overview of the map when using the last-seen image representation. Using the arrow, seems to be more difficult compared to using the map. Participants stated that they did not know if they were heading towards the right room e.g. if the object was behind a wall and they needed to switch rooms. During the study we observed that the participants were faster when they were navigating with the map than when they were using the arrow. On room level, the participants were faster when they were using the last-seen image than when they were using the map.. Therefore, we propose a hybrid system using the map for finding the correct room and using the last-seen image for finding the sought object. In addition, the arrow caused problems due to the fact that the compass can be easily irritated inside a building. Compass irritations happen irregularly, but often (every 1 to 2 minutes in average). This was one of the major reasons why we decided to conduct the study in a wizard-of-oz manner (see Section 5.1.3 for details). The proposed system could therefore be used without the need for a wizard-of-oz.

Interestingly, our results are partly in conflict with the results shown in the preliminary study. In the preliminary study users voted the map as the most favored representation on room-level. In contrast, our results showed the last-seen image as the most favored representation on room-level. However, our results showed the map to be the most favored representation in general. This is probably due to the fact that most participants disliked the arrow for finding the correct room. The remaining findings support the results of the preliminary study, in which we compared textual, map, 3D, and last-seen image representation on room level. The differences of the SUS and TLX values are statistically insignificant.

The Google Maps Navigation bears an analogy to our proposed design. Using the Google Maps Navigation, a user is shown a map to navigate to a destination. When reaching the destination, an image is shown (most likely from Google Street view). The image shows the destination.

However, there is a factor which needs to be considered in evaluating the findings of the study. All objects were hidden in an obvious way. No object lay in any container, like a drawer or a locker, and was not covered in any way. The only object that was hidden a little harder is the one we can show significant differences in the task completion time. Nevertheless, it's position was still trivial, on a stool next to a shelf. The object was not covered in any way. Participants who did not use the last-seen image just passed by the object. Probably because the object was not in their field of sight. In everyday life most irretrievable objects are hidden hard or at least unconventional. In some drawer covered by paperwork or in a shelf behind other objects. A future study should include harder hidden objects to be more realistic and therefore determine if the task completion time using the last-seen image is significantly faster than using the map. The current findings indicate that the last-seen image is faster and less mentally demanding for hard search tasks. However, no throughout statistical significant results could be shown.

The study shows that, since all objects have been found, both object location representations generally work in a large scale scenario. One object has been found faster using the last-seen image representation. The other objects were not found faster using either one of the representations. Subjective user feedback showed that people liked the map to plan their route

towards the sought object. The last-seen image was preferred for object retrieval on room-level. Therefore, we recommend using the map to find the correct room and using the last-seen image on room-level.

6 Conclusion and Future Work

In this thesis we explored a novel object location representation on building-level: the last-seen image. The last-seen image is produced as a byproduct of a real-world search engine using visual object recognition. The design space of the last-seen image is underexplored. To shed some light on the usability of the last-seen image representation we conducted a user study to evaluate the last-seen image compared to a map based representation.

We compared two object location representations on building-level. The hypotheses was that using the last-seen image is more efficient than using the map representation. We were able to show that both representations work properly throughout a building consisting of several floors. The hypotheses was not fully confirmed even though the findings support the hypotheses. It was shown that the last-seen image works better for hard hidden objects. The experimental setup should nevertheless be improved in a followup study. Our experimental setup consisted of trivially hidden objects and thus users were not challenged.

The user provided feedback hints to use the map as the room finding technique. The last-seen image is favored in terms of object retrieval on room-level. Therefore, we propose a hybrid system using the map for finding the correct room and using the last-seen image for finding the sought object.

Overall, the last-seen image is a model-independent object location representation working out of the box. A real-world search engine that only uses the last-seen image is suitable in a small scaled environment. The visual object recognition can be done on the HMD. No server or any additional setup is required. No calibration or model is needed as well as no WiFi indoor positioning. Thus, a real world search engine using visual object recognition is feasible with todays technology. That is the power of the last-seen image.

6.1 Future Work

The main focus of this thesis is on evaluating the last-seen image. We therefore decided to touch only some topics of the concept and design space of a real-world search engine using visual object recognition. However, in the future we want to provide a fully functional system that works out of the box. A lot of design decisions have to be taken according to the used technologies. If we focus on small scaled environments like homes or small offices, we will not need WiFi indoor positioning. Relying exclusively on the last-seen image would be sufficient in

that case. In large scaled environments like a campus or a company site indoor positioning is mandatory.

6.1.1 Followup study using Harder Hidden Objects

A major limitation of the presented user study is that the objects were hidden trivially. All objects were visible when entering the room in which they lay. Our results hint that the last-seen image becomes significantly better than the map when used on hard hidden objects.

Therefore, we propose an experimental design using harder hidden objects. In reality, lost objects that cannot be retrieved immediately are not visible when entering the room. We recommend hiding all or a subset of all objects inside of drawers, covered by other objects, inside of lockers, under a couch, and so on. We believe using such a setup will show significantly better results for the last-seen image compared to the map. The residual setup may remain the same (see Chapter 5.1 for more information).

6.1.2 Visual Object Recognition

The focus of this thesis lies on the comparison and exploration of object location representations. To be able to provide a fully working real-world search engine the object recognition has to be fully functional. Visual object recognition using a wearable camera has already been explored [Funk et al., 2013].

We propose a setup using a centralized database that at least consists of the last-seen image and coordinates of the last seen position of each object. The visual object recognition engine detects an object and streams the last-seen image to the database. A user searching for an object can query the database using the wearable system proposed in the system description (see Chapter 4).

6.1.3 Future Work on the Design Space and Concept

We were not able to do a whole exploration of the design space, but we provide advice for future attempts that should be considered in our opinion. In addition, we also provide future work suggestions regarding real-world search engines using visual object recognition. As stated before, this was not part of the thesis and therefore no further evaluation has been performed. Thus, each topic is only touched.

Last-Seen Image

We showed the usability of the last-seen image throughout this thesis. A short discussion on how to capture good last-seen images and how to rate them among each other has been carried out (see Chapter 3.5.5). Most objects are not constantly moved, but are rather static. Thus, after some time there will be a lot of look alike last-seen images showing the same picture from different angles and distances. Consequently, some images provide more helpful contextual information than others. An algorithm deciding which last-seen image shall be shown to the user should be evaluated.

Handling Lookalike Objects

Lookalike objects are recognized as only one object. Therefore, a way to distinguish these objects is needed . We proposed possible solutions in the design space (see Chapter 3.5.3).

Outdoor Usage

Scenarios in an outside environment are thinkable as well. For instance finding the parked car (see Chapter 3.4.2).

Bibliography

SeungJun Kim and Anind K. Dey. Simulated augmented reality windshield display as a cognitive mapping aid for elder driver navigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 133–142, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-246-7. doi:10.1145/1518701.1518724. URL http://doi.acm.org/10.1145/1518701.1518724. (Cited on page 7)

Wolfgang Narzt, Gustav Pomberger, Alois Ferscha, Dieter Kolb, Reiner Müller, Jan Wieghardt, Horst Hörtner, and Christopher Lindinger. Augmented reality navigation systems. *Univers. Access Inf. Soc.*, 4(3):177–187, February 2006. ISSN 1615-5289. doi:10.1007/s10209-005-0017-5. URL http://dx.doi.org/10.1007/s10209-005-0017-5. (Cited on page 7)

Gerhard Reitmayr and Dieter Schmalstieg. Collaborative augmented reality for outdoor navigation and information browsing. In *Proceedings of the Second Symposium on Location Based Services and TeleCartography*, pages 53–62. TU Wien, 2004. URL http://publik.tuwien.ac.at/files/PubDat_137965.pdf. (Cited on pages 7 and 18)

David Kirsh. The intelligent use of space. *Artif. Intell.*, 73(1-2):31–68, February 1995. ISSN 0004-3702. doi:10.1016/0004-3702(94)00017-U. URL http://dx.doi.org/10.1016/0004-3702(94)00017-U. (Cited on page 8)

Esure. We're a bunch of 'losers'. *http://www.esure.com/media_centre/archive/wcmcap_100800.html*, 2012. (Cited on page 8)

Rodney E Peters, Richard Pak, Gregory D Abowd, Arthur D Fisk, and Wendy A Rogers. Finding lost objects: Informing the design of ubiquitous computing services for the home. *GVU Technical Report;GIT-GVU-04-01*, 2004. URL http://hdl.handle.net/1853/51. (Cited on page 8)

Giuseppe Amato, Fabrizio Falchi, and Paolo Bolettieri. Recognizing landmarks using automated classification techniques: Evaluation of various visual features. In *Proceedings of the 2010 Second International Conferences on Advances in Multimedia*, MMEDIA '10, pages 78–83, Washington, DC, USA, 2010. IEEE Computer Society. ISBN 978-0-7695-4068-9. doi:10.1109/MMEDIA.2010.20. URL http://dx.doi.org/10.1109/MMEDIA.2010.20. (Cited on page 8)

Robin Boldt, Marcus Eisele, and Yalcin Taha. Vergleich der Visualisierungsmöglichkeiten einer Real-World Search-Engine, 2013. (Cited on pages 9, 19, 21, 22, 30 and 43)

Markus Funk, Albrecht Schmidt, and Lars Erik Holmquist. Antonius: A mobile search engine for the physical world. In *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication*, UbiComp '13 Adjunct, pages 179–182, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2215-7. doi:10.1145/2494091.2494149. URL http://doi.acm.org/10.1145/2494091.2494149. (Cited on pages 10, 15, 21, 29 and 54)

Markus Funk, Robin Boldt, Bastian Pfleging, Max Pfeiffer, Niels Henze, and Albrecht Schmidt. Representing indoor location of objects on wearable computers with head-mounted displays. In *Proceedings of the 5th Augmented Human International Conference*, AH '14, pages 18:1–18:4, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2761-9. doi:10.1145/2582051.2582069. URL http://doi.acm.org/10.1145/2582051.2582069. (Cited on pages 11 and 77)

Julie A. Kientz, Shwetak N. Patel, Arwa Z. Tyebkhan, Brian Gane, Jennifer Wiley, and Gregory D. Abowd. Where's my stuff?: Design and evaluation of a mobile system for locating lost items for the visually impaired. In *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '06, pages 103–110, New York, NY, USA, 2006. ACM. ISBN 1-59593-290-9. doi:10.1145/1168987.1169006. URL http://doi.acm.org/10.1145/1168987.1169006. (Cited on page 13)

Christian Frank, Philipp Bolliger, Christof Roduner, and Wolfgang Kellerer. Objects calling home: Locating objects using mobile phones. In *Proceedings of the 5th International Conference on Pervasive Computing*, PERVASIVE'07, pages 351–368, Berlin, Heidelberg, 2007. Springer-Verlag. ISBN 978-3-540-72036-2. URL http://dl.acm.org/citation.cfm?id=1758156.1758183. (Cited on page 13)

J. M. Kahn, R. H. Katz, and K. S. J. Pister. Next century challenges: Mobile networking for “smart dust”. In *Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking*, MobiCom '99, pages 271–278, New York, NY, USA, 1999. ACM. ISBN 1-58113-142-9. doi:10.1145/313451.313558. URL http://doi.acm.org/10.1145/313451.313558. (Cited on page 14)

Thomas Pederson. Magic touch: A simple object location tracking system enabling the development of physical-virtual artefacts in office environments. *Personal Ubiquitous Comput.*, 5(1):54–57, January 2001. ISSN 1617-4909. doi:10.1007/s007790170031. URL http://dx.doi.org/10.1007/s007790170031. (Cited on pages 14 and 28)

Mizuho Komatsuzaki, Koji Tsukada, Itiro Siio, Pertti Verronen, Mika Luimula, and Sakari Pieskä. Iteminder: Finding items in a room using passive rfid tags and an autonomous robot (poster). In *Proceedings of the 13th International Conference on Ubiquitous Computing*, UbiComp '11, pages 599–600, New York, NY, USA, 2011a. ACM. ISBN 978-1-4503-0630-0. doi:10.1145/2030112.2030232. URL http://doi.acm.org/10.1145/2030112.2030232. (Cited on page 14)

Lionel M. Ni, Yunhao Liu, Yiu Cho Lau, and Abhishek P. Patil. Landmarc: Indoor location sensing using active rfid. *Wirel. Netw.*, 10(6):701–710, November 2004. ISSN 1022-

0038. doi:10.1023/B:WINE.0000044029.06344.dd. URL http://dx.doi.org/10.1023/B:WINE.0000044029.06344.dd. (Cited on pages 14 and 19)

Kok-KIONG Yap, Vikram Srinivasan, and Mehul Motani. Max: Wide area human-centric search of the physical world. *ACM Trans. Sen. Netw.*, 4(4):26:1–26:34, September 2008. ISSN 1550-4859. doi:10.1145/1387663.1387672. URL http://doi.acm.org/10.1145/1387663.1387672. (Cited on page 14)

Jens Nickels, Pascal Knierim, Bastian Könings, Florian Schaub, Björn Wiedersheim, Steffen Musiol, and Michael Weber. Find my stuff: Supporting physical objects search with relative positioning. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '13, pages 325–334, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1770-2. doi:10.1145/2493432.2493447. URL http://doi.acm.org/10.1145/2493432.2493447. (Cited on page 14)

Andreas Butz, Michael Schneider, and Mira Spassova. Searchlight – a lightweight search function for pervasive environments. In Alois Ferscha and Friedemann Mattern, editors, *Pervasive Computing*, volume 3001 of *Lecture Notes in Computer Science*, pages 351–356. Springer Berlin Heidelberg, 2004. ISBN 978-3-540-21835-7. doi:10.1007/978-3-540-24646-6_26. URL http://dx.doi.org/10.1007/978-3-540-24646-6_26. (Cited on page 15)

Peter Ljungstrand, Johan Redström, and Lars Erik Holmquist. Webstickers: Using physical tokens to access, manage and share bookmarks to the web. In *Proceedings of DARE 2000 on Designing Augmented Reality Environments*, DARE '00, pages 23–31, New York, NY, USA, 2000. ACM. doi:10.1145/354666.354669. URL http://doi.acm.org/10.1145/354666.354669. (Cited on page 15)

Jun Rekimoto and Yuji Ayatsuka. Cybercode: Designing augmented reality environments with visual tags. In *Proceedings of DARE 2000 on Designing Augmented Reality Environments*, DARE '00, pages 1–10, New York, NY, USA, 2000. ACM. doi:10.1145/354666.354667. URL http://doi.acm.org/10.1145/354666.354667. (Cited on page 15)

Mizuho Komatsuzaki, Koji Tsukada, and Itiro Siio. Drawerfinder: Finding items in storage boxes using pictures and visual markers. In *Proceedings of the 16th International Conference on Intelligent User Interfaces*, IUI '11, pages 363–366, New York, NY, USA, 2011b. ACM. ISBN 978-1-4503-0419-1. doi:10.1145/1943403.1943466. URL http://doi.acm.org/10.1145/1943403.1943466. (Cited on pages 15 and 28)

Roy Want, Andy Hopper, Veronica Falcão, and Jonathan Gibbons. The active badge location system. *ACM Trans. Inf. Syst.*, 10(1):91–102, January 1992. ISSN 1046-8188. doi:10.1145/128756.128759. URL http://doi.acm.org/10.1145/128756.128759. (Cited on page 16)

Andy Harter, Andy Hopper, Pete Steggles, Andy Ward, and Paul Webster. The anatomy of a context-aware application. In *Proceedings of the 5th Annual ACM/IEEE International*

Conference on Mobile Computing and Networking, MobiCom '99, pages 59–68, New York, NY, USA, 1999. ACM. ISBN 1-58113-142-9. doi:10.1145/313451.313476. URL http://doi.acm.org/10.1145/313451.313476. (Cited on page 16)

Nissanka B. Priyantha, Anit Chakraborty, and Hari Balakrishnan. The cricket location-support system. In *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking*, MobiCom '00, pages 32–43, New York, NY, USA, 2000. ACM. ISBN 1-58113-197-6. doi:10.1145/345910.345917. URL http://doi.acm.org/10.1145/345910.345917. (Cited on page 16)

P. Bahl and V.N. Padmanabhan. Radar: an in-building rf-based user location and tracking system. In *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 775–784 vol.2, 2000. doi:10.1109/INFCOM.2000.832252. (Cited on page 16)

Robert J. Orr and Gregory D. Abowd. The smart floor: A mechanism for natural user identification and tracking. In *CHI '00 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '00, pages 275–276, New York, NY, USA, 2000. ACM. ISBN 1-58113-248-4. doi:10.1145/633292.633453. URL http://doi.acm.org/10.1145/633292.633453. (Cited on page 16)

Juhi Ranjan, Yu Yao, and Kamin Whitehouse. An rf doormat for tracking people's room locations. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '13, pages 797–800, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1770-2. doi:10.1145/2493432.2493514. URL http://doi.acm.org/10.1145/2493432.2493514. (Cited on page 17)

Oliver Woodman and Robert Harle. Pedestrian localisation for indoor environments. In *Proceedings of the 10th International Conference on Ubiquitous Computing*, UbiComp '08, pages 114–123, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-136-1. doi:10.1145/1409635.1409651. URL http://doi.acm.org/10.1145/1409635.1409651. (Cited on pages 17 and 22)

Alessandro Mulloni, Daniel Wagner, Istvan Barakonyi, and Dieter Schmalstieg. Indoor positioning and navigation with camera phones. *IEEE Pervasive Computing*, 8(2):22–31, April 2009. ISSN 1536-1268. doi:10.1109/MPRV.2009.30. URL http://dx.doi.org/10.1109/MPRV.2009.30. (Cited on page 17)

M. Kalkusch, T. Lidy, Michael Knapp, G. Reitmayr, H. Kaufmann, and D. Schmalstieg. Structured visual markers for indoor pathfinding. In *Augmented Reality Toolkit, The First IEEE International Workshop*, pages 8 pp.–, 2002. doi:10.1109/ART.2002.1107018. (Cited on page 17)

G. Schroth, R. Huitl, D. Chen, M. Abu-Alqumsan, A. Al-Nuaimi, and E. Steinbach. Mobile visual location recognition. *Signal Processing Magazine, IEEE*, 28(4):77–89, July 2011. ISSN 1053-5888. doi:10.1109/MSP.2011.940882. (Cited on page 17)

Jaewoo Chung, Matt Donahoe, Chris Schmandt, Ig-Jae Kim, Pedram Razavai, and Micaela Wiseman. Indoor location sensing using geo-magnetism. In *Proceedings of the 9th international conference on Mobile systems, applications, and services*, MobiSys '11, pages 141–154, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0643-0. doi:10.1145/1999995.2000010. URL http://doi.acm.org/10.1145/1999995.2000010. (Cited on page 17)

Christian Kray, Christian Elting, Katri Laakso, and Volker Coors. Presenting route instructions on mobile devices. In *Proceedings of the 8th International Conference on Intelligent User Interfaces*, IUI '03, pages 117–124, New York, NY, USA, 2003. ACM. ISBN 1-58113-586-6. doi:10.1145/604045.604066. URL http://doi.acm.org/10.1145/604045.604066. (Cited on page 17)

Benjamin Walther-Franks and Rainer Malaka. Evaluation of an augmented photograph-based pedestrian navigation system. In *Proceedings of the 9th International Symposium on Smart Graphics*, SG '08, pages 94–105, Berlin, Heidelberg, 2008. Springer-Verlag. ISBN 978-3-540-85410-4. doi:10.1007/978-3-540-85412-8_9. URL http://dx.doi.org/10.1007/978-3-540-85412-8_9. (Cited on page 18)

Patrick Baudisch and Ruth Rosenholtz. Halo: A technique for visualizing off-screen objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 481–488, New York, NY, USA, 2003. ACM. ISBN 1-58113-630-7. doi:10.1145/642611.642695. URL http://doi.acm.org/10.1145/642611.642695. (Cited on page 18)

Antonio Krüger, Andreas Butz, Christian Müller, Christoph Stahl, Rainer Wasinger, Karl-Ernst Steinberg, and Andreas Dirschl. The connected user interface: Realizing a personal situated navigation service. In *Proceedings of the 9th International Conference on Intelligent User Interfaces*, IUI '04, pages 161–168, New York, NY, USA, 2004. ACM. ISBN 1-58113-815-6. doi:10.1145/964442.964473. URL http://doi.acm.org/10.1145/964442.964473. (Cited on page 18)

L. Ran, S. Helal, and S. Moore. Drishti: an integrated indoor/outdoor blind navigation system and service. In *Pervasive Computing and Communications, 2004. PerCom 2004. Proceedings of the Second IEEE Annual Conference on*, pages 23–30, March 2004. doi:10.1109/PERCOM.2004.1276842. (Cited on page 18)

J.A.B. Link, P. Smith, N. Viol, and K. Wehrle. Footpath: Accurate map-based indoor navigation using smartphones. In *Indoor Positioning and Indoor Navigation (IPIN), 2011 International Conference on*, pages 1–8, Sept 2011. doi:10.1109/IPIN.2011.6071934. (Cited on page 18)

Andreas Butz, Jörg Baus, Antonio Krüger, and Marco Lohse. A hybrid indoor navigation system. In *Proceedings of the 6th International Conference on Intelligent User Interfaces*, IUI '01, pages 25–32, New York, NY, USA, 2001. ACM. ISBN 1-58113-325-1. doi:10.1145/359784.359832. URL http://doi.acm.org/10.1145/359784.359832. (Cited on page 18)

Alessandro Mulloni, Hartmut Seichter, and Dieter Schmalstieg. Handheld augmented reality indoor navigation with activity-based instructions. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '11, pages 211–220, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0541-9. doi:10.1145/2037373.2037406. URL http://doi.acm.org/10.1145/2037373.2037406. (Cited on page 18)

Andreas Möller, Matthias Kranz, Stefan Diewald, Luis Roalter, Robert Huitl, Tobias Stockinger, Marion Koelle, and Patrick Lindemann. Experimental evaluation of user interfaces for visual indoor navigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 1–10, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2473-1. doi:10.1145/2556288.2557003. (Cited on page 18)

Frank Biocca, Arthur Tang, Charles Owen, and Fan Xiao. Attention funnel: Omnidirectional 3d cursor for mobile augmented reality platforms. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 1115–1122, New York, NY, USA, 2006. ACM. ISBN 1-59593-372-7. doi:10.1145/1124772.1124939. URL http://doi.acm.org/10.1145/1124772.1124939. (Cited on page 19)

Kevin Ashton. That 'Internet of Things' Thing. *RFID Journal*, 2011. (Cited on page 21)

Stephen Se, David Lowe, and Jim Little. Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. *The international Journal of robotics Research*, 21(8): 735–758, 2002. (Cited on page 23)

Shyang-Lih Chang, Li-Shien Chen, Yun-Chung Chung, and Sei-Wan Chen. Automatic license plate recognition. *Intelligent Transportation Systems, IEEE Transactions on*, 5(1):42–53, March 2004. ISSN 1524-9050. doi:10.1109/TITS.2004.825086. (Cited on page 25)

Allyson Rice, P Jonathon Phillips, Vaidehi Natu, Xiaobo An, and Alice J O'Toole. Unaware person recognition from the body when face identification fails. *Psychological science*, 24 (11):2235–2243, 2013. (Cited on page 27)

Markus Funk. Searching the real world using stationary and mobile object detection. Master's thesis, Universität Stuttgart, Holzgartenstr. 16, 70174 Stuttgart, 2012. URL http://elib.uni-stuttgart.de/opus/volltexte/2013/8793. (Cited on pages 27, 28 and 33)

Werner Vogels. Eventually consistent. *Commun. ACM*, 52(1):40–44, January 2009. ISSN 0001-0782. doi:10.1145/1435417.1435432. URL http://doi.acm.org/10.1145/1435417.1435432. (Cited on page 28)

Niels Henze, Torben Schinke, and Susanne Boll. What is that? object recognition from natural features on a mobile phone. In *Proceedings of the Workshop on Mobile Interaction with the Real World*, pages 101–112. Citeseer, 2009. (Cited on page 29)

Roger W Sinnott. Virtues of the haversine. *Sky and telescope*, 68:158, 1984. (Cited on page 38)

J.P. Snyder. *Flattening the Earth: Two Thousand Years of Map Projections*. University of Chicago Press, 1997. ISBN 9780226767475. URL http://books.google.de/books?id=0UzjTJ4w9yEC. (Cited on page 38)

Chris Veness. Calculate distance and bearing between two latitude/longitude points using haversine formula in javascript, 2012. *http://www.movable-type.co.uk/scripts/latlong.html*, 2012. (Cited on page 39)

E. Vonesh and V.M. Chinchilli. *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. Statistics: A Series of Textbooks and Monographs. Taylor & Francis, 1996. ISBN 9780824782481. URL http://books.google.de/books?id=SK_Er5_tmAgC. (Cited on page 44)

Rosemary A Bailey. *Design of Comparative Experiments*, volume 25. Cambridge University Press, 2008. ISBN 9781139469913. URL http://books.google.de/books?id=mvSHP6Cbx8sC. (Cited on page 44)

S. G. Hart and L. E. Stavenland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, pages 139–183, 1988. URL http://humansystems.arc.nasa.gov/groups/tlx/downloads/NASA-TLXChapter.pdf. (Cited on page 44)

John Brooke. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189:194, 1996. (Cited on page 44)

Sandra G. Hart. Nasa-Task Load Index (Nasa-TLX); 20 Years Later. In *Human Factors and Ergonomics Society Annual Meeting*, volume 50, 2006. URL http://humansystems.arc.nasa.gov/groups/tlx/downloads/HFES_2006_Paper.pdf. (Cited on page 44)

N. Dahlback, A. Jonsson, and L. Ahrenberg. Wizard of Oz-studies – why and how. In *Workshop on Intelligent User Interfaces*, Orlando, FL, 1993. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.56.3398. (Cited on page 45)

American Psychological Association. *Publication manual of the American Psychological Association*. American Psychological Assoc., Washington, DC, 2012. ISBN 1433805596 1433805618 1433805626 9781433805592 9781433805615 9781433805622. URL http://www.worldcat.org/search?qt=worldcat_org_all&q=1433805618. (Cited on page 47)

All links were last followed on May 1, 2014.

Appendix

The appendix contains the forms, as handed out to the participants, of the beforehand described study (see Chapter 5). We provide the forms in the order they were shown to the participants. The first appendix is the guideline that was used by the experimenter. Followed by the consent form, demographics, system usability scale (SUS), task load index (TLX), and final questions. Some of the appendix is in German since the experiment took place in Germany. However, we used the English version of the tests (TLX and SUS). In our opinion the existing official translations are insufficient. We used Google Forms[1] in order to create the questionnaires.

[1]https://support.google.com/drive/answer/87809?hl=en

Studien Leitfaden 07.01.2014

Einführung:

- Die Studie ist Teil meiner Bachelorarbeit.
- Wir wollen die Repräsentationsarten einer tragbaren Real World Search Engine evaluieren.
- Was ist eine Real World Search Engine ?
    - Vergleichbar mit der Suche nach Informationen im Internet - Suchmaschine z.B. Google
    - Es werden Informationen zu Gegenständen in der echten Welt gesucht - z.B. wo ist mein Schlüssel.
- Wir wollen zwei Repräsentationsarten zur Gegenstandslokalisierung vergleichen.
- Eine Suche wird folgendermaßen ablaufen:
    - Nennung des Gegenstandes durch den Studienleiter und zeigen des Reference Images.
    - Nach dem Gegenstand suchen (Starte Stoppuhr beim ersten Schritt des Probanden).
    - Den Gegenstand antippen, dann gilt er als gefunden (Stoppe Zeitmessung).
- Während der Durchführung unbedingt auf die persönliche Sicherheit achten!
- Consent Form sowie Teilnehmernummer aufüllen lassen.
- Den Proband mit der Brille und Suchengine vertraut machen.
    - Suche nach Gegenständen.
    - Mute-Knopf erklären.

Durchführung:

- Zu Beginn jeder Repräsentationsart, wird eine Testrunde durchgeführt (Kaffeemaschine bei Karte, Feuerlöscher bei Bild).
- Blau Position - Rot Ziel
- Jede Condition startet vor Raum 01.025.
- Jede Condition beinhaltet 3 Gegenstände. Jeweils einen Small, Medium, Large.
- Nach jeder Condition ein SUS und TLX ausfüllen lassen.

Auswertung:

- Aufteilung der Fragen nach Bild/Karte.
    - Was fandest du gut?
    - Was fandest du schlecht?
    - Was könnte man verbessern?
- In welchen Situationen würdest du so ein System einsetzen wollen? Warum?
- Wie müsste das System aussehen, damit du es nutzen würdest?

University of Stuttgart
Germany

Human Computer Interaction Group (MCI), VIS

Prof. Dr. Albrecht Schmidt

Consent Form

DESCRIPTION: You are invited to participate in **a research study** on **object retrieval in the context of a wearable real world search engine**.

TIME INVOLVEMENT: Your participation will take approximately **45 minutes.**

DATA COLLECTION: For this study you will be asked to retrieve objects with the help of a wearable device. You will need to fill in questionnaires. Also, you will be interviewed at the end of the study.

RISKS AND BENEFITS: There is the risk of a collision, while you walk through the building. Never risk an accident. There is no need to hurry. The collected data is securely stored. We do guarantee no data misuse and privacy is completely preserved. Your decision whether or not to participate in this study will not affect your grade in school.

PAYMENTS: You will receive **sweets** as reimbursement for your participation.

PARTICIPANT'S RIGHTS: If you have read this form and have decided to participate in this project, please understand your **participation is voluntary** and you have the **right to withdraw your consent or discontinue participation at any time without penalty or loss of benefits to which you are otherwise entitled**. **The alternative is not to participate.** You have the right to refuse to answer particular questions. The results of this research study may be presented at scientific or professional meetings or published in scientific journals. Your identity is not disclosed unless we directly inform and ask for your permission.

CONTACT INFORMATION: If you have any questions, concerns or complaints about this research, its procedures, risks and benefits, contact following persons:
Robin Boldt (boldtrn@gmail.com)
Markus Funk (markus.funk@vis.uni-stuttgart.de).

By signing this document I confirm that I agree to the terms and conditions.

Name: _____ *Signature, Date:* _____

Teilnehmernummer

* Required

Teilnehmernummer *

Alter *

Geschlecht *

Studiengang/Beruflicher Hintergrund *

Wie oft suchen Sie durchschnittlich nach Gegenständen? *

Wie z.B nach Autoschlüsseln, Handy, Kamera etc.

Submit

Never submit passwords through Google Forms.

System Usability Scale

* Required

Teilnehmernummer *

Repräsenationsart *

Map ▼

I think that I would like to use this system frequently: *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

I found the system unnecessarily complex: *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

I thought the system was easy to use: *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

I think that I would need the support of a technical person to be able to use this system: *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

I found the various functions in this system were well integrated: *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

I thought there was too much inconsistency in this system: *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

I would imagine that most people would learn to use this system very quickly: *

	1	2	3	4	5

Strongly disagree ○ ○ ○ ○ ○ Strongly agree

I found the system very cumbersome to use: *

1 2 3 4 5

Strongly disagree ○ ○ ○ ○ ○ Strongly agree

I felt very confident using the system: *

1 2 3 4 5

Strongly disagree ○ ○ ○ ○ ○ Strongly agree

I needed to learn a lot of things before I could get going with this system: *

1 2 3 4 5

Strongly disagree ○ ○ ○ ○ ○ Strongly agree

Submit

Never submit passwords through Google Forms.

Edit this form

TLX

* Required

Teilnehmernummer *

Repräsentationsart *

Scales *

_____Very Low _____Very High

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Mental Demand - How mentally demanding was the task?	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Physical Demand - How physically demanding was the task?	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Temporal Demand - How hurried or rushed was the pace of the task?	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Performance - How successful were you in accomplishing what you were asked to do?	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Effort - How hard did you have to work to accomplish your level of performance?	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Frustration - How insecure, discouraged, irritated, stressed, and annoyed wereyou?	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

Submit

Never submit passwords through Google Forms.

Edit this form

FinalQuestions

* Required

Teilnehmernummer *

Fragen zur Kartendarstellung

Was fandest du gut?

Was fandest du schlecht?

Was würdest du verbessern?

Fragen zum Last Seen Image

Was fandest du gut?

Was fandest du schlecht?

Was würdest du verbessern?

Allgemeine Fragen

In welchen Situationen würdest du so ein System einsetzen wollen?

Warum?

Wie müsste das System aussehen, damit du es nutzen würdest?

Welche ist deine favorisierte Darstellungsart? *

Submit

Never submit passwords through Google Forms.

Declaration

I hereby declare that the work presented in this thesis is entirely my own and that I did not use any other sources and references than the listed ones. I have marked all direct or indirect statements from other sources contained therein as quotations. Neither this work nor significant parts of it were part of another examination procedure. Parts of this thesis have been published as research paper at the 5th Augmented Human International Conference 2014 in Kobe, Japan. The title of this publication is “Representing Indoor Location of Objects on Wearable Computers with Head-Mounted Displays” [Funk et al., 2014]. The electronic copy is consistent with all submitted copies.

---

place, date, signature