A Mobile Indoor Navigation System Interface Adapted to Vision-Based Localization

Andreas Möller
Technische Universität München
Institute for Media Technology
Munich, Germany
andreas.moeller@tum.de

Matthias Kranz
Luleå University of Technology
Department of
Computer Science,
Electrical and Space Engineering
Luleå, Sweden
matthias.kranz@ltu.se

Robert Huitl, Stefan Diewald, Luis Roalter
Technische Universität München
Institute for Media Technology
Munich, Germany
huitl@tum.de,
stefan.diewald@tum.de,
roalter@tum.de

ABSTRACT

Vision-based approaches for mobile indoor localization do not rely on the infrastructure and are therefore scalable and cheap. The particular requirements to a navigation user interface for a vision-based system, however, have not been investigated so far.

Such mobile interfaces should adapt to localization accuracy, which strongly relies on distinctive reference images, and other factors, such as the phone's pose. If necessary, the system should motivate the user to point at distinctive regions with the smartphone to improve localization quality.

We present a combined interface of Virtual Reality (VR) and Augmented Reality (AR) elements with indicators that help to communicate and ensure localization accuracy. In an evaluation with 81 participants, we found that AR was preferred in case of reliable localization, but with VR, navigation instructions were perceived more accurate in case of localization and orientation errors. The additional indicators showed a potential for making users choose distinctive reference images for reliable localization.

Categories and Subject Descriptors

H.5.m [**Information Interfaces and Presentation**]: Miscellaneous

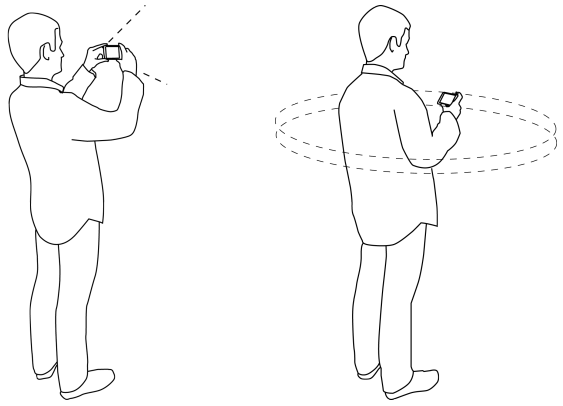
General Terms

Human Factors, Design.

Keywords

Vision-based localization, augmented reality, virtual reality, user interface, indoor navigation

MUM '12, December 04–06 2012, Ulm, Germany



(a) Augmented reality (AR) shows guidance as overlays on real-time video

(b) Virtual reality (VR) guides the user in a pre-recorded panorama view

Figure 1: We propose an augmented reality or virtual reality visualization depending on how the user carries the phone (upright or down) and on the location estimate's accuracy in order to improve the user experience and perceived quality of navigation instructions.

1. INTRODUCTION

As pedestrian outdoor navigation has become ubiquitous through GPS-enabled smartphones, the demand for reliable localization indoors as well is significantly increasing. Indoor localization and navigation is considered an enabler for a variety of applications, such as guidance of passengers on airports, conference attendees, visitors in shopping malls, hospitals or office buildings, and for many novel context-aware services, which can play a significant role for monetarization.

The traditionally used outdoor localization method GPS is not available in indoor environments. A catalog of alternative localization techniques has been investigated, such as Infrared- [6], sensor- [34] and radio-based technologies [17] or visual markers [29]. All those technologies, however, need a particular environmental infrastructure and augmentation.

Image matching using feature extraction (e.g. [9]) can, in some contexts, be a promising alternative, especially for large-scale environments. With this technique, query images, captured by the localization system, are matched with

reference images based on discriminative descriptors (so-called features). The known location of the most similar reference image is then used as location estimate. Scale- and rotation-invariant features (e.g. [22]) facilitate matching even if the reference image was recorded from a different angle or position. Today's high-quality cameras in smartphones make them ready for employing vision-based localization. We see a potential in this technique especially as it does not impose any infrastructural requirements.

While a huge body of research exists for the technical background of feature-based localization [33, 4, 7, 31, 12], the investigation of the human-computer interaction (HCI) perspective thereof is just at its beginnings. Previous research so far has mostly focused on algorithms that have been evaluated with sample data sets. Real-world conditions impose challenges that have not been considered so far. For example, uniform indoor spaces such as long corridors or motion blur in the query images can lead to insufficient or nondiscriminative features. The resulting inaccuracy of the location and orientation estimate is challenging for the user interface.

In this paper, we examine the intersection of vision-based localization with HCI. We argue that the user interface for a mobile indoor navigation system that relies on image matching needs to reflect and address the particularities of this technique. While various indoor navigation interfaces have been designed so far (we will provide an overview in the related work section), none have been deliberately conceived for vision-based localization.

Our contributions are twofold: First, we present a thorough concept for a user interface suited for vision-based indoor navigation. Second, we provide results of an extensive evaluation of our concept with 81 subjects, based on video simulations and mock-ups. In particular, we address the ability to deal with location and orientation inaccuracy and report on subjects' perception and feedback.

The paper is structured as follows. We start by giving an overview on related work and summarize the advantages and challenges of vision-based localization. After that, we introduce and explain our user interface concept and components. Subsequently, the evaluation is presented and the results are discussed in light of a future implementation of the system.

2. BACKGROUND AND RELATED WORK

Mobile indoor navigation systems have been in focus of research recently (for a survey, see e.g. [13]). Kray et al. [18] provide an overview of how route instructions can be presented on mobile devices. They distinguish different classes of interfaces, such as textual and spoken instructions, 2D sketches, 2D maps and pseudorealistic 3D views. Each of them can be classified according to the adequacy for different levels of localization accuracy. For example, a location marker in a map requires position, but no orientation information, and a pre-rendered 3D animation of the route can be used without location or orientation information. 3D was preferred over 2D in a user study due to its inclusion of landmarks that helped subjects locate themselves on the route [18].

Technical capabilities, like the quality of the location estimate or the abilities of a device constrain information presentation (such as 3D graphics). Beyond that, also the cognitive resources of the user play a role for the choice of the interface [18]. A map view can e.g. be rotated with the user's orientation for less cognitive effort with self-localization [18, 6]. Two-dimensional maps can be adapted to the user's walking speed [5]. If the user walks faster, the map zooms out to show more of the environment. The inaccuracy of the location estimate is illustrated with a circle around the position marker on the map [6, 5]. The system by Butz et al. [6] switches between different visualizations based on quality of the location and orientation estimate. While in case of high accuracy a simple arrow is sufficient, a map view and landmarks, such as elevators and staircases, are included with decreasing accuracy. The authors argue that a high resolution of location and orientation information is not always required. With an orientation resolution of 45 degrees, the system can give correct instructions on a T-junction. Likewise, the system does not have to know one's exact location in a long corridor; it can be sufficient to distinguish between two decision points.

Augmented reality (AR), i.e. the integration of virtual objects in a real-world scene [2], has been used in manifold ways for navigation user interfaces, e.g. based on visual markers [15] or using image-based localization [23]. The strength of AR is its intuitiveness, since no translation between the virtual representation and the real world is required [30]. A survey of AR systems and applications is given e.g. by Azuma et al. [3]. Narzt et al. [30] used AR to visualize paths for car navigation in style of a head-up display. These paths go beyond directional arrows and also e.g. illustrate exits that are hidden behind a truck and include safety aspects such as highlighting people crossing the road. Their concept also encompasses context-based notifications of e.g. prices when passing by a gas station. For indoor use, Liu et al. [19] presented a system for the smartphone that uses superimposed directional arrows in combination with textual navigation instructions and audio. They found that the ability to adapt the interface for users' preferences is particularly important. Miyashita et al. [25] used AR for a museum guidance system. Augmentations were used to enhance exhibits with additional information. At the same time visitors were guided along a predefined route through the museum when they searched with their phone for the next AR object. In a study by Walther-Franks and Malaka [32], subjects evaluated the usability of an AR navigation system better than map-based system. The system used floor-projected arrows and lines as way directions.

Representing an environment with omnidirectional (i.e., panoramic) images [21, 8] reduces the graphical effort to create a full 3D scene. Their photorealism makes panoramas well-suited for navigation systems, since surrounding elements can serve as orientation landmarks. Miyazaki et al. [26] present such a system where panoramas are generated on a server and sent to the handheld client device. Outdoor approaches where image material is augmented with directional information are presented for example in [16] and [11]. A similar indoor approach is shown by Merico and Bisiani [24] using a gyroscope and dead reckoning for localization.

With our interface concept, we focus on vision-based localization using a smartphone camera (for a survey of other techniques suitable for indoor use, see [20]). Such approaches have been presented e.g. by [10], [29], and [33]. In Section 3, we motivate the application of vision techniques for local-

ization in indoor environments, but also address the concomitant challenges. In our work, however, we focus on the requirements to the user interface, not the underlying techniques.

3. ADVANTAGES AND CHALLENGES OF VISION-BASED LOCALIZATION

Vision as a basis for indoor localization provides a number of advantages, but also challenges, in particular with relation to the user interface which we explore in the following.

3.1 Advantages

Other active localization methods (i.e., where the device localizes itself actively) rely on an augmented environment, such as electronic tags or a WLAN access point infrastructure. This is difficult to handle in large-scale environments, since the augmentation does not scale well and is expensive. Additionally, no generally accepted standard exists for indoor localization systems (e.g. with respect to technology, method, and associated parameters, such as accuracy).

Vision-based localization in this regard provides several advantages. First, it only relies on a camera (and thus only has requirements in the visual domain) and sufficient processing power, which is both given in up-to-date smartphones (in sufficient quality and resolution). Additionally, vision-based localization can be combined with inertial sensors [1] such as accelerometers, which are likewise built into state-of-the art phones. Hence, current hardware already fulfills these requirements. Second, the fact that augmentation of the environment is unnecessary simplifies the process of deploying and using such a system. It reduces costs and is feasible in arbitrary surroundings. Augmentation (e.g. with beacons) is not possible everywhere due to constructional challenging environments, energy consumption (power supply is not given everywhere), vandalism, costs or legal problems. Moreover, vision-based localization is suited, in principle, for indoor and outdoor environments and a seamless transition between them.

3.2 Challenges

Localization using vision entails, however, also some challenges. First, it requires reference data, i.e. the environment must be known in order to localize the device within the environment. Reference images must be gathered in the first place and the exact location must be assigned to each image, e.g. by using a mapping trolley as presented in [14]. Since the environment could be subject to change (e.g. when shop window displays, adverts or posters are replaced), a way to update the reference material must be foreseen. This can be done centralized or in a collaborative approach, where query material is tagged with a location manually by users and eventually becomes part of the new reference dataset.

A second challenge is the quality and distinctiveness of the query images, which impact the location estimate. Motion might make the camera-visible scene blurry. Moreover, the typical pose (i.e., orientation) when holding a phone (about 45 degrees downwards) entails that corridors and halls (being good candidates for reference images) are not visible to the camera. As a consequence, not enough visual features can be extracted for reliable localization. In contrast to GPS- or radio-based localization systems, where the device's orientation does usually not play a role, vision-based localization requires the camera to point at 'interesting' regions in the environment. The ideal pose therefore would be as if taking a photo. Permanently maintaining this pose is inconvenient for the user. We thus need to ensure that location accuracy remains sufficient for providing correct navigation instructions, until the next accurate localization is possible when the camera sees enough visual features again (correctness here means application-dependent accuracy). It is the role of user interface here to take necessary actions if required, while ensuring convenient navigation feedback. Since the pose of the smartphone is not fixed (unlike when mounted in a car), the interface has to adapt to changing situations, such as the usage while walking (quick glance to check the correct way) or while standing (re-orientation or exploring). Sometimes it has to persuade the user to change the pose of the device to ensure a desired level of localization accuracy. We report in the following on our approach towards achieving this goal.

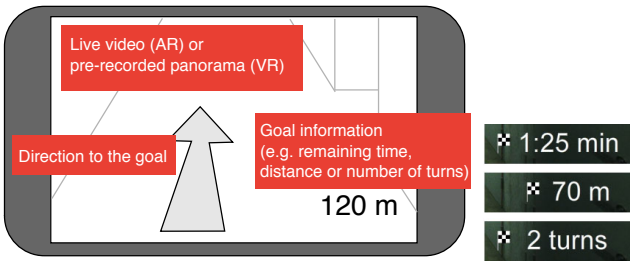


Figure 2: The user interface of our proposed concept consists of a perspectively displayed navigation arrow which is included in a pre-rendered panorama of the environment (virtual reality) or imposed over the live video (augmented reality). In the bottom right, additional information, such as the remaining time, distance or number of turns to the goal, can be displayed (see screenshots from the mockup).

4. A USER INTERFACE ADAPTED TO VISION-BASED INDOOR USE

In this work, we propose and evaluate a user interface concept for a smartphone indoor navigation system [28] that addresses the challenges of vision-based localization introduced above. The interface incorporates three key components which are described in the following: 1) augmented and virtual reality, 2) indicators that communicate and ensure accuracy, and 3) area of interest indicators.

4.1 Augmented and Virtual Reality

In the related work section, we gave an overview on various approaches for pedestrian navigation. A considerable part thereof used some sort of augmented reality to visualize navigation instructions. However, there are several reasons against using *only* augmented reality overlays. They always require an upright pose of the smartphone where the user 'looks through' to see both the environment and the augmented overlays. This pose is inconvenient for long-term or frequent use (e.g. in unknown environments), so that an alternative visualization is required which can also be employed when the user is looking down. The user will also not always hold the device in a way that the camera sees

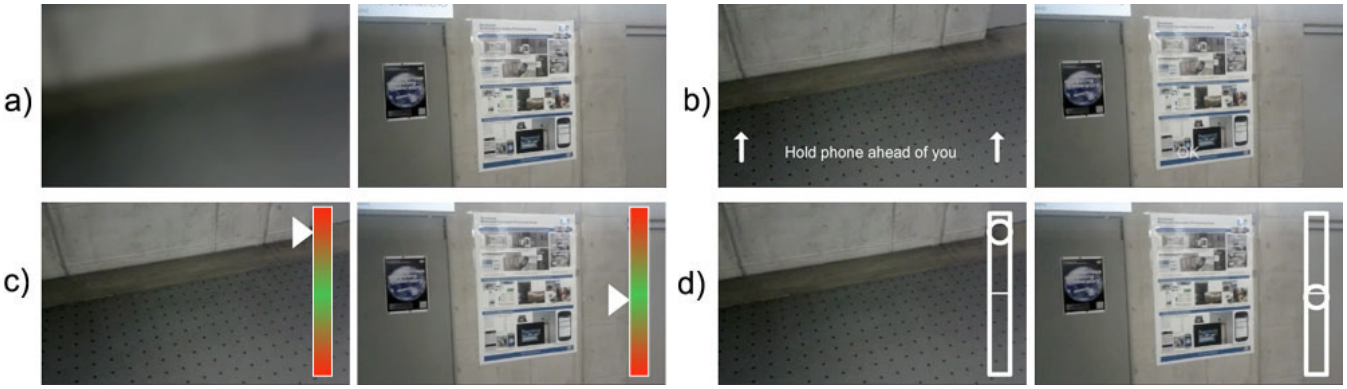


Figure 3: Proposed visualizations to raise the smartphone to discriminative image areas in eye height, leading to more reliable localization. a) Blur metaphor (focus change), b) text instruction, c) color scale, d) water level metaphor. The left images of each visualization show the view when the smartphone points downwards, right images after it has been directed to a feature-rich area.

enough discriminative features. If the location estimate consequently becomes inaccurate, the alternative visualization should still be able to present reliable navigation instructions. For that reason, we propose two alternative visualizations that can be used for indoor navigation.

- **Augmented Reality (AR).** This visualization augments the video seen by the smartphone's camera by superimposing navigation information, such as a directional arrow and the distance to the next turn. Users needs to hold the phone upright as illustrated in Fig. 1(a) in order to see the augmentation directly on their way. The video image is used to localize the device in the environment (in terms of position and orientation), so that the overlays can be accurately placed.
- **Virtual Reality (VR).** This visualization uses prerecorded images of the environment that are stitched to a 360 degree panorama on the mobile device. Since reference images are required anyway for vision-based localization, the material is readily available. The advantage of VR is that the user does not have to hold the smartphone up as if looking 'through' the phone to see navigation instructions. Instead, directional arrows are directly embedded in the panorama, so that the device can be held in a more natural and comfortable way (see Fig. 1(b)). Due to the known geometry, the arrow can be rendered more optimally in the panorama than in AR. The best-matching panorama image is retrieved based on the current location estimate. The panorama can manually be dragged around by the user for self-orientation or be rotated automatically with the determined orientation.

Figure 2 shows a mockup of the proposed navigation system that could either represent the AR or VR system. In case of AR, the arrow is imposed on the live video image and changes with the location estimate. In case of VR, the hallway image is prerecorded and the orientation of the arrow is fixed in relation to the panorama image.

4.2 Communicating and Ensuring Accuracy

In order to additionally ensure the technical goal of a lower bound for the quality of the location estimate, we propose an additional user interface element which ensures that sufficient visual features can be detected by the system. An indicator appears in case of low localization quality, prompting the user to actively point at regions containing more visual features. In light of the fact that distinctive, feature-rich areas are typically found in eye sight (e.g. door signs, posters, showcases, advertisements), this could typically be achieved by raising the phone up, to a pose as depicted in Figure 1(a).

- **Blur.** In analogy to a camera focusing on the motive, artificial focus change is used to guide the user towards a feature-rich area. Starting from a blurry scene, the image gets sharper as the user approaches a feature-rich area (see Fig. 3(a)).
- **Text.** A simple text hint is displayed, indicating to move the smartphone in a specific direction (see Fig. 3(b)).
- **Color Scale.** A color-coded scale ranging from red (bottom/top, symbolizing few features) to green (center, symbolizing enough features) represents the number of distinctive visual features in the image, and changes with the user pointing towards a non-uniform area (see Fig. 3(c)).
- **Water Level.** The metaphor of a water level is used to indicate the correct orientation of the phone. For an optimal position, the vial should be aligned in the center of the level (see Fig. 3(d)).

4.3 Highlighting Interesting Areas

Augmented reality browsers, such as Junaio[1] are nowadays very popular. Hence, the integration of AR elements *beyond* navigation instructions suggests itself for a vision-based localization system. Highlighting objects around the user can be an enabler for context-based services. For example, shops in a mall, special offers, individual shop windows or even doors and doorplates can be points of interaction. Using object recognition techniques (cf. [27] for a recent comparison of state-of-the-art techniques), interactive objects can be tracked without additional markers and highlighted to the user when the device is held as depicted in Figure 1(a).

[1]http://www.junaio.com

Figure 4: Highlighting of interaction points for context-based services. Instead of framing objects (left), we propose a soft border visualization (right) for less sensitivity to jitter and to reduce distraction of the user.

However, if points of interest are permanently visible during the navigation task, users could be distracted by the additional overlays and visual elements present on the screen. In addition, inherent tracking inaccuracy can make the overlay jitter, which could further distract the user. We propose a visualization using a soft border that smoothly fades out around the object. We hypothesize that inherent inaccuracies could be better hidden due to the lack of a defined border, adding to a more stable, calm visualization. Figure 4 illustrates the difference between a conventional frame object highlighting and the soft border visualization.

As a side effect, such areas of interaction are presumably rich in distinctive visual features. If they attract the user's attention and are focused with the smartphone's camera, they implicitly serve for improving the system's certainty of the location estimate.

5. EVALUATION

In the following, we present the research questions regarding our proposed user interface concept, describe the proceeding of the evaluation and discuss the results.

5.1 Research Questions

We investigated the following research questions.

RQ1. Which concept (AR or VR) is preferable in terms of perceived accuracy?
We want to investigate whether and how the visualization influences the perceived accuracy, related to position and orientation, and the quality of navigation instructions. We hypothesize that VR, where the navigation arrow has a fixed direction in relation to the panorama, can improve the impression of accuracy. Especially when the system's location or orientation estimate is wrong, we expect that the perceived reliability of the system is increased in VR, compared to AR.

RQ2. Which concept (AR or VR) is preferred by users?
We investigate subjects' preferences for a particular visualization, which need to be taken into account as well for a convenient user experience.

RQ3. What information should be presented?
Additional information, such as the remaining time, distance or number of turns to the goal, can be presented to the user.

Figure 5: A screenshot from one of the videos used in the study, showing the simulated field of view (top) and the navigation system mockup (bottom).

Since the requirements to indoor navigation differ from traditional outdoor or car navigation, it has to be investigated what pieces of information are considered as important under those special conditions.

RQ4. Which visualizations could be appropriate to acquire sufficient visual features?
Vision-based localization needs distinctive features to function. These are often found in eye height, and thus only in reach of the smartphone's camera if the user holds it upright. We investigate which visualizations intuitively motivate the user to raise the phone, in order to assist the system in gaining sufficient features for reliably localizing the device.

RQ5. Can object highlighting be improved with a soft border visualization?
For context-based services, objects the user can interact with must be highlighted. However, unstable object tracking can lead to jiggling visualizations that are liable to irritate the user. We investigate if the newly presented technique using soft borders (see Section 4.3) leads to improved perception.

5.2 Research Method and Approach

Our goal was to gain initial insights and answers to the above research questions. Envisaging an iterative design process, we wanted to collect feedback from a large number of users before we would implement our proposed concepts or revise them according to our findings. The evaluation was conducted based on mock-up images and videos, illustrating the operation of the system, and a corresponding online questionnaire. Subjects were asked to watch the video demonstrations or look at the mockup images, respectively, and to answer questions related to the presented material.

In order to have subjects estimate how they perceive accuracy in the AR and VR system, we prepared a pre-recorded sample navigation route. The video with a duration of about 45 seconds was played back alongside with the simulated output of the system. The video demonstrations contained the simulated field of view (i.e., the 'reality') in the upper part, and the simulated visualization on the smartphone in the lower part. Fig. 5 shows a screenshot of one of the videos used in the study. For both AR and VR, we systematically introduced artificial errors to the system's location estimate in terms of location and orientation. The navigation instruc-

tions varied then based on these errors, so that they showed for example wrongly oriented arrows, or loaded mismatching panoramas. We used four conditions (no error, location error, orientation error, combined error) with four different types of errors.

- **No Error.** All navigation instructions were correct.
- **Location Error.** An error was introduced as it would occur when the system's estimated location was wrong. This type of error manifests in panorama images of a wrong location (for VR), or wrong turn instructions (for AR, e.g. when the system thinks of being next to a corridor where there is none). This error was induced twice in the "location error" condition.
- **Orientation Error.** An error was introduced as it would occur when the system's estimated orientation was wrong. This type of error manifests in incorrectly rotated panorama images (for VR), or incorrectly rotated arrows (for AR). This error was induced twice in the "orientation error" condition.
- **Combined Error.** Both location and orientation errors were introduced twice in this condition.

All subjects ran through all conditions (within-subjects design, four videos each for AR and VR). The order of conditions was counter-balanced using a latin square design; subjects did not know which error condition they were currently evaluating when watching the videos.

Similarly, the feature indicators (*Blur*, *Text*, *Color*, *Water Level*) were presented to participants in four videos in permuted order using a latin square in order to exclude potential learning effects.

5.3 Participants

Participants were recruited using the Mobileworks crowdsourcing platform[2]. 81 subjects, aged between 18 and 59 years (average age: 28, standard deviation = 8.7), participated in the study; 39 thereof were female, 42 were male.

For a representative user basis, we did not require particular familiarity or experience with navigation systems. 43% of our subjects indicated to use navigation systems infrequently (several times a month), 18% use them often (several times a week) and 4% very often (daily). 35% never use navigation systems at all. 40% declared to be experienced with car navigation systems, 26% had used pedestrian navigation before, and 12 % stated to have experience with an indoor navigation system.

5.4 Results

All answers were given on 7-step Likert scales ranging from -3 (strongly disagree) to +3 (strongly agree). For all ratings, the standard deviation is provided (in the following abbreviated with SD). If not explicitly otherwise stated, all comparative results between conditions were statistically significant ($p < 0.001$ in a Student's t-test).

5.4.1 RQ 1: Perceived Accuracy of AR and VR

Augmented Reality. Fig. 6 summarizes the evaluation of AR and VR with relation to the perceived accuracy.

In the *no error* condition, subjects felt that the system knew well their location (2.5, SD = 0.9) and orientation (2.4, SD = 1.0). As expected, this perceived accuracy significantly decreased in the error conditions. With an *orientation error*, subjects answered on average only with 0.8 (SD = 2.0) that the system was certain about their location, and with 0.2 (SD = 2.1) that it was sure about their orientation. For *location errors*, ratings were 1.7 (SD = 1.5) for the perceived location accuracy and 1.2 (SD = 1.8) for the perceived orientation accuracy. The perceived accuracy further decreased for the combined error condition. Here, the rating was 0.6 (SD = 2.0) for location accuracy and 0.4 (SD = 2.1) for orientation accuracy.

Participants perceived the correctness of the navigation instructions as follows: In the *no error* condition, they averagely rated the correctness with 2.3 (SD = 1.0). With orientation and location errors, this rating decreased to -0.2 (SD = 2.0) and 0.4 (SD = 1.9), and with both error types together to -0.5 (SD = 1.9).

The results show a significant decrease of the perceived accuracy between the *no error* and the single error conditions (*orientation* and *location*) with $p < 0.001$ in the Student's t-test. The perceived accuracy in the *combined error* condition was significantly lower than in the *location error* condition. The difference between the *combined error* and the *orientation error* condition was not statistically significant ($p > 0.1$).

In the single error conditions, the perceived accuracy of location and orientation decreased both. In the *orientation error* condition, subjects perceived the orientation correctly as less accurate than the location. However, in the *location error* condition, subjects had the same impression, although here the orientation would have been expected to be more accurate than then location. This indicates that subjects had problems to distinguish location and orientation errors. In fact, both may lead to wrongly orientated navigation arrow overlays. Not only an orientation estimation error may cause the navigation arrow overlay to point in a wrong direction, but also if the system locates the user further away from a crossing than she actually is (location error), the arrow may point in a wrong direction (even backwards).

Virtual Reality. In the *no error* condition, subjects evaluated the perceived location and orientation estimate's accuracy with 1.7 (SD = 1.6 and 1.5). With the introduced orientation error, the rating slightly decreased to 1.4 (SD = 1.8) for the location estimate and to 1.1 (SD = 1.8) for the orientation estimate. In the *location error* condition, the perceived accuracy decreased to 1.4 (SD = 1.8) for the location estimate and 1.3 (SD = 1.7) for the orientation estimate. When both errors were combined, the perceived accuracy was rated with 1.0 (SD = 1.7) for location and with 0.9 (SD = 1.8) for orientation. The perceived correctness of navigation instructions decreased from 1.8 (SD = 1.4) in the *no-error* condition to 1.4 in the single-error conditions (SD = 1.6 for orientation error and 1.7 for location error), and to 0.9 (SD = 1.8) in the *combined error* condition.

The decreases between the *no error* and the single-error conditions (*location* or *orientation*) are partly significant ($p < 0.05$ for perceived orientation and correctness of navigation instructions, but not for perceived location). The differences are highly significant between the *no error* and *combined error* conditions ($p < 0.001$). There was a significant decrease in perceived accuracy between the *location* and the *combined error* condition ($p < 0.05$), but not be-

[2]www.mobileworks.com

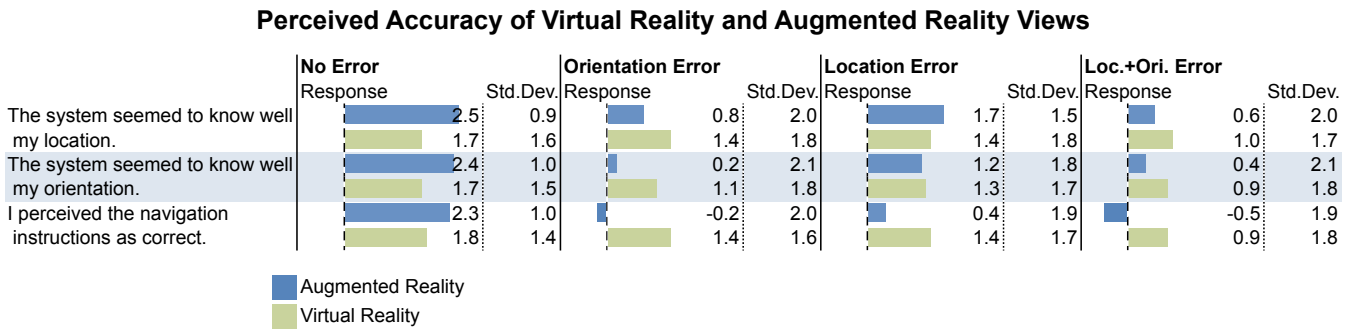


Figure 6: Perceived accuracy of virtual and augmented reality visualizations (agreement to statements on a 7-step Likert scale; -3=strongly disagree, 3=strongly agree). Std.Dev. denotes the standard deviation.

tween the *orientation* and the *combined error* condition ($p > 0.1$). Although most differences are statistically significant, the perceived accuracies of location and orientation remain fairly high throughout all error conditions (compared to AR results).

Comparison of AR and VR. In VR, the ratings throughout all conditions were more similar than in AR. AR was perceived as more accurate than VR in case of a perfectly working system, i.e., in the *no error* condition ($p < 0.001$ for position/location estimates and $p < 0.05$ for perceived correctness of navigation instructions). However, in the error conditions, subjects felt that VR was more reliable. No significant differences of the perceived accuracy between AR and VR in the *location* condition were found, but subjects perceived VR to be more accurate ($p < 0.05$) in the *orientation error* and *combined error* conditions. The navigation instructions were perceived significantly more correct with VR throughout all error conditions ($p < 0.001$).

5.4.2 RQ 2: User Preferences for AR or VR

Immediately after having experienced either the AR or the VR visualization, subjects were asked for their personal opinion of the respective system (see Fig. 7 for a summary of the results. In these questions, subjects showed a significant tendency towards the AR system ($p < 0.05$ in Student's t-test). Asked whether they could imagine to use either one of the systems themselves, they agreed on average with 1.9 (SD = 1.3) that they could imagine using AR, but only agreed on average with 1.1 (SD = 1.8) for VR. The high standard deviation in case of VR shows that subjects were controversial in that point. In a direct vote which system they liked more, subjects clearly favored AR (58%). VR was chosen by 24%, and 18% were undecided.

5.4.3 RQ 3: Goal Information

We asked which additional information about the goal (the remaining distance, the number of turns until the goal, or the remaining time) was considered as important (see Fig. 2, screenshots on the right). Subjects found the distance information most useful (averagely rated with 2.1, SD = 1.3), followed by the time (1.3, SD = 1.8) and by the number of turns (1.1, SD = 1.8). The differences are highly significant ($p < 0.001$ in a Student's t-test), except for the difference between time and number of turns. We also presented variants where each of the three pieces of information

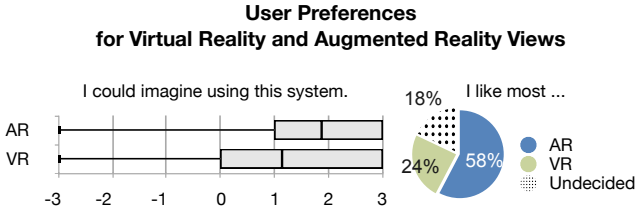


Figure 7: User preferences for the virtual and augmented reality system. Subjects preferred augmented reality (AR) over virtual reality (VR) navigation instructions in the evaluation. Answers given on a 7-step Likert scale; -3=strongly disagree, 3=strongly agree.

was enhanced with a bar that illustrated the ratio of elapsed and remaining distance/turns/time graphically. There was no significantly different voting for the variants with bar, so that this could be an optional setting up to individual preferences.

5.4.4 RQ 4: Feature-Rich Area Indicators

Subjects were presented four indicators (*Text*, *Water Level*, *Color* and *Blur*, see Section 4.2) that intend to make the user direct the phone to feature-rich regions. We evaluated the intuitiveness of the presented visualizations. The results are summarized in Fig. 8. Best results were obtained for the *Text* and the *Water Level* visualization. Subjects responded that the meaning was clear on average with 1.7 (SD = 1.5) for *Text*, and that of *Water Level* with 1.5 (SD = 1.6). *Color* was evaluated with 1.1 (SD = 1.7). The intuitiveness of *Blur* was below average and showed a high standard deviation (-0.2, SD = 2.2). A Student's t-test showed that the difference between *Text* and *Water Gauge* was not significant ($p > 0.1$), but all other differences were statistically significant with $p < 0.05$.

5.4.5 RQ 5: Object of Interest Indicators

We evaluated the visualization for highlighting objects of interest as presented in Section 4.3 (in the following called *Soft Border*) and compared it against a conventional highlighting method, where a frame around the recognized object is used (in the following called *Frame*). Subjects stated that both highlighting methods almost equally raised their

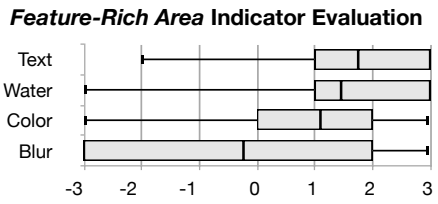


Figure 8: User ratings of the clarity of the indicators' meaning. The Text and Water Gauge visualizations were evaluated significantly better than a color scale or a blur/focus metaphor. Answers given on a 7-step Likert scale; -3=strongly disagree, 3=strongly agree.

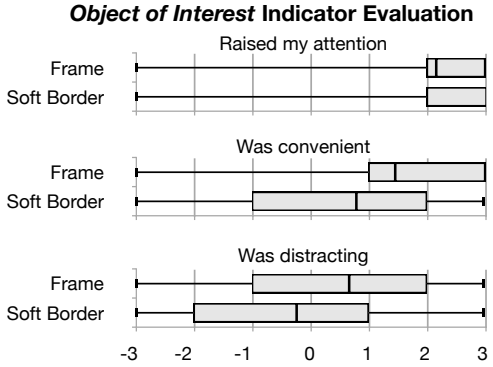


Figure 9: User ratings of different highlighting and tracking visualizations of interesting objects. At similar level of attention, a soft border highlight (cf. Fig. 4) was perceived as less distracting than a border around the object, but subjects also found it less convenient. Answers given on a 7-step Likert scale; -3=strongly disagree, 3=strongly agree.

attention to the object. They agreed that their attention was directed towards the visualization on average with 2.2 for *Frame* (SD = 1.3) and with 2.0 for *Soft Border* (SD = 1.4). The difference was not significant ($p > 0.1$ in a Student's t-test). The visualizations were evaluated significantly different ($p < 0.05$) related to their convenience. The convenience of the *Frame* visualization was evaluated with 1.5 (SD = 1.6), while *Soft Border* was only rated with 0.8 (SD = 1.9). *Soft Border* was found significantly less distracting in their navigation task ($p < 0.05$) than *Frame*: for *Soft Border*, subjects agreed on average with -0.2 (SD = 2.0) that this visualization would distract them, while the rating for *Frame* was 0.7 (SD = 1.9).

5.4.6 Additional Findings

For some visualizations, we considered minor variants that we presented to our subjects as well for evaluation.

We provided different versions of the VR system where we modified the frequency in which panoramas were replaced. The system used for comparison against the AR system updated the panorama about every second. In addition, we provided a version with faster update rate (every 0.5 seconds) and with slower update rate (every 2 seconds). The one-second frequency was appreciated slightly more (average 1.0, SD = 1.7) than faster (0.8, SD = 1.8) or slower panorama changes (0.6, SD = 1.7). Only the difference between medium and slow transitions was significant (t-test with $p < 0.05$).

We also varied the way how panorama changes were performed. We showed a version with hard changes (which was used in the comparison against AR), one with soft transitions where one panorama dissolved in the subsequent one, and a version using a zoom animation, blending from one perspective to the next. Here, soft transitions were evaluated slightly better with 0.6 (SD = 2.1) than hard transitions (0.4, SD = 1.9) and zooming (0.4, SD = 1.9), but differences were not significant.

For each of the object-of-interest indicators (*Frame* and *Soft Border*), we presented a version where the background video was in color, ane one desaturated version with black and white video (b/w). We hypothesized that the b/w version could further focus attention to the object and thus be beneficial. Results showed however no significant differences between color and b/w backgrounds. In Sec. 5.4.5 we have described the indicators with colored background. The b/w variants were evaluated with 2.0 (SD = 1.5) for *Frame*, and with 2.0 (SD = 1.4) for *Soft Border*. This is a difference of 0.2 points on the Likert scale for *Frame* and 0.0 points for *Border*, so that this effect can most probably be neglected.

6. DISCUSSION

Perception of Accuracy Through the Interface

Our proposed navigation interface concepts, Augmented Reality and Virtual Reality, show their strengths in different domains. In case of incorrect location and orientation estimates, VR was perceived as more reliable, since it was less influenced by inaccuracies. Users can match panoramic images with the environment also if they are slightly translated or rotated to the actual position.

By contrast, AR instructions imposed on live video appear wrong and misleading if the estimated position is erroneous. A wrongly estimated orientation was perceived more negatively than a wrongly estimated location. In an implemented version, the reliability of orientation estimates could therefore be increased by a combination of the vision-based estimation with device sensors (e.g., a compass). Since the certainty of the localization estimate in most systems can easily be calculated, an automated choice of the situationally optimal visualization (AR or VR) would be possible.

Of course, the system should communicate ambiguous self-localization estimates, in particular when those errors can lead to wrong navigation instructions, so that the user can choose the right path by himself. In situations where a lower accuracy is sufficient (e.g. in a long corridor without junctions), VR can hide inherent inaccuracies better and therefore increase trust in a navigation system. However, in case of reliable localization, AR was generally perceived as more accurate than VR, and it was also the preferred visualization by users when asked directly. Therefore, we argue that a combination of both systems could be beneficial, or even necessary, for vision-based navigation. The considerable standard deviations for in particular the AR ratings reflect that users are heterogeneous and do not perceive the same. This must be taken into account as well in an implemented system.

Ensuring Accuracy With Good Reference Images
Our analysis of *feature indicators* addressed the important point of creating awareness for how well a scene serves for localization and how the user can assist the system to improve accuracy. Sufficient salient features in the image are crucial for reliable vision-based localization. In this study, we investigated suitable metaphors and showed in a first step that a visualization like the water level metaphor, possibly in combination with text instructions, is intuitive and understandable. In a next step, it will have to be investigated in a real-world study whether such visualizations are actually an incentive to focus on feature-rich areas in eye height.

Situational Use of VR and AR
The process of acquiring visual features also correlates with the choice of a VR or AR visualization. VR is rather suited when holding the phone in a 45 degree angle, as depicted in Fig. 1(b). The user can then compare the panoramic image on the phone with his field of view. By contrast, AR only makes sense when the phone is held upright, as illustrated in Fig. 1(a). In this mode, the user sees the environment 'through' the phone. The AR pose is also required for targeting visual features which is a further argument why a vision-based system should not solely rely on a VR interface (the VR interface provides no incentive for holding the phone upright).

This observation is a further advantage of a combined and situation-based AR and VR interface. When the phone is carried normally the camera points towards the floor and usually few visual features are in sight of the camera. The system will then have to rely on relative positioning based on the latest absolute location estimate. Here, VR can be used (which does not require that much accuracy for still working satisfactory). When accuracy falls below a threshold, the user needs to lift up the phone and the AR interface is activated. Once sufficient visual features are captured, the location estimate can be updated.

Context-Based Services
Object highlighting and tracking can provide hints to the user which objects are augmented with additional information, such as posters, showcases, doors, or devices (elevator controls, photocopiers, etc.). They are important for context-based services, but should at the same time not distract the user during the navigation task. Although we have evaluated all interface parts individually in these study, they are combined in a final implementation.

Results show that our method of soft borders equally draws attention to the highlighted object, but reduces distraction and thus might interfere less with the navigation task. The actual effect under real-world conditions will have to be investigated in future work. A manual turn on/off solution could be a way to enable highlights on a user-based preference. Distraction could further be reduced by defining categories or priorities of allowed highlights during navigation. For example, the navigation goal could be allowed to be highlighted, but not other objects around the user.

7. CONCLUSION AND FUTURE WORK

The results of our study provide first insights on the evaluated individual user interface parts. We believe that a combination of AR and VR is indeed adequate for indoor navigation, and for the particularities of vision-based localization. While some results were highly significant, others showed high variances and reflect that user preferences are heterogeneous. In practice, they can also depend on individual goals. As a major next step, a state model could therefore be deduced that determines which action should take place in which state of the system, and which visualization should accordingly be chosen. The choice of AR or VR will most likely be dependent on the phone's pose (upright or down) and the current location estimate's certainty, and intelligently trigger notifications for pose changes if necessary. It will thereby be important that the user is included in the loop in a discreet way. Required user interaction should not only have the purpose to help the system to function, but also match the user's own intentions and goals. The highlighting and presentation of ambient objects for location-based services that we have presented could be one possibility towards this goal. Those extensions, and real-world studies of the proposed concepts, are subject to future work.

Acknowledgments

This research project has been supported by the space agency of the German Aerospace Center with funds from the Federal Ministry of Economics and Technology on the basis of a resolution of the German Bundestag under the reference 50NA1107.

8. REFERENCES

[1] M. Angermann and P. Robertson. Footslam: Pedestrian simultaneous localization and mapping without exteroceptive sensors—hitchhiking on human perception and cognition. *Proceedings of the IEEE*, 100(13):1840–1848, 2012.

[2] R. Azuma. A survey of augmented reality. *Presence-Teleoperators and Virtual Environments, MIT Press*, 6(4):355–385, 1997.

[3] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre. Recent advances in augmented reality. *Computer Graphics and Applications, IEEE*, 21(6):34–47, 2001.

[4] G. Baatz, K. Köser, D. Chen, R. Grzeszczuk, and M. Pollefeys. Leveraging 3D city models for rotation invariant place-of-interest recognition. *International Journal of Computer Vision*, 96(3):315–334, 2012.

[5] J. Baus, A. Krüger, and W. Wahlster. A resource-adaptive mobile navigation system. In *Proceedings of the 7th international conference on Intelligent user interfaces*, pages 15–22. ACM, 2002.

[6] A. Butz, J. Baus, A. Krüger, and M. Lohse. A hybrid indoor navigation system. In *Proceedings of the 6th international conference on Intelligent user interfaces*, pages 25–32. ACM, 2001.

[7] D. Chen, G. Baatz, K. Köser, S. Tsai, R. Vedantham, T. Pylvanainen, K. Roimela, X. Chen, J. Bach, M. Pollefeys, B. Girod, and R. Grzeszczuk. City-scale landmark identification on mobile devices. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 737–744, Colorado Springs, USA, 2011.

[8] S. Chen. Quicktime VR: An image-based approach to virtual environment navigation. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 29–38. ACM, 1995.

[9] G. Fritz, C. Seifert, and L. Paletta. A mobile vision system for urban detection with informative local descriptors. In *IEEE International Conference on Computer Vision Systems (ICVS'06)*, pages 30–30. IEEE, 2006.

[10] H. Hile and G. Borriello. Positioning and orientation in indoor environments using camera phones. *Computer Graphics and Applications, IEEE*, 28(4):32–39, 2008.

[11] H. Hile, R. Vedantham, G. Cuellar, A. Liu, N. Gelfand, R. Grzeszczuk, and G. Borriello. Landmark-based pedestrian navigation from collections of geotagged photos. In *Proceedings of the 7th International Conference on Mobile and Ubiquitous Multimedia*, pages 145–152. ACM, 2008.

[12] S. Hilsenbeck, A. Möller, R. Huitl, G. Schroth, M. Kranz, and E. Steinbach. Scale-preserving long-term visual odometry for indoor navigation. In *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2012.

[13] H. Huang and G. Gartner. A survey of mobile indoor navigation systems. *Cartography in Central and Eastern Europe*, pages 305–319, 2010.

[14] R. Huitl, G. Schroth, S. Hilsenbeck, F. Schweiger, and E. Steinbach. TUMindoor: An extensive image and point cloud dataset for visual indoor localization and mapping. In *IEEE ICIP*, Miami, USA, 2012.

[15] M. Kalkusch, T. Lidy, N. 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. IEEE, 2002.

[16] T. Kolbe. Augmented videos and panoramas for pedestrian navigation. In *Geowissenschaftliche Mitteilungen*. Citeseer, 2003.

[17] M. Kranz, C. Fischer, and A. Schmidt. A comparative study of DECT and WLAN signals for indoor localization. In *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 235–243. IEEE, 2010.

[18] C. Kray, C. Elting, K. Laakso, and V. Coors. Presenting route instructions on mobile devices. In *Proceedings of the 8th international conference on Intelligent user interfaces*, pages 117–124. ACM, 2003.

[19] A. Liu, H. Hile, H. Kautz, G. Borriello, P. Brown, M. Harniss, and K. Johnson. Indoor wayfinding: Developing a functional interface for individuals with cognitive impairments. *Disability & Rehabilitation: Assistive Technology*, 3(1-2):69–81, 2008.

[20] H. Liu, H. Darabi, P. Banerjee, and J. Liu. Survey of wireless indoor positioning techniques and systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 37(6):1067–1080, 2007.

[21] P. Liu, X. Sun, N. Georganas, and E. Dubois. Augmented reality: a novel approach for navigating in panorama-based virtual environments (PBVE). In *2nd IEEE International Workshop on Haptic, Audio and Visual Environments and Their Applications (HAVE)*, pages 13–18. IEEE, 2003.

[22] D. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, volume 2, pages 1150–1157. IEEE, 1999.

[23] J. Menzel, M. Königs, and L. Kobbelt. A framework for vision-based mobile ar applications. In *Proceedings of the Workshop on Mobile Vision and HCI (MobiVis). Held in Conjunction with Mobile HCI*, 2012.

[24] D. Merico and R. Bisiani. Indoor navigation with minimal infrastructure. In *4th Workshop on Positioning, Navigation and Communication (WPNC'07)*, pages 141–144. IEEE, 2007.

[25] T. Miyashita, P. Meier, T. Tachikawa, S. Orlic, T. Eble, V. Scholz, A. Gapel, O. Gerl, S. Arnaudov, and S. Lieberknecht. An augmented reality museum guide. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 103–106. IEEE Computer Society, 2008.

[26] Y. Miyazaki and T. Kamiya. Pedestrian navigation system for mobile phones using panoramic landscape images. In *International Symposium on Applications and the Internet (SAINT 2006)*. IEEE, 2006.

[27] A. Möller, S. Diewald, L. Roalter, and M. Kranz. MobiMed: Comparing Object Identification Techniques on Smartphones. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction (NordiCHI 2012)*, pages 31–40, Copenhagen, Denmark, 2012. ACM.

[28] A. Möller, C. Kray, L. Roalter, S. Diewald, R. Huitl, and M. Kranz. Tool Support for Prototyping Interfaces for Vision-Based Indoor Navigation. In *Proceedings of the Workshop on Mobile Vision and HCI (MobiVis). Held in Conjunction with Mobile HCI*, 2012.

[29] A. Mulloni, D. Wagner, I. Barakonyi, and D. Schmalstieg. Indoor positioning and navigation with camera phones. *Pervasive Computing, IEEE*, 8(2):22–31, 2009.

[30] W. Narzt, G. Pomberger, A. Ferscha, D. Kolb, R. Müller, J. Wieghardt, H. Hörtner, and C. Lindinger. Augmented reality navigation systems. *Universal Access in the Information Society*, 4(3):177–187, 2006.

[31] G. Schroth, R. Huitl, D. Chen, M. Abu-Alqumsan, A. Al-Nuaimi, and E. Steinbach. Mobile visual location recognition. *IEEE Signal Processing Magazine*, 28(4):77–89, 2011.

[32] B. Walther-Franks and R. Malaka. Evaluation of an augmented photograph-based pedestrian navigation system. In *Smart Graphics*, pages 94–105. Springer, 2008.

[33] M. Werner, M. Kessel, and C. Marouane. Indoor positioning using smartphone camera. In *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–6. IEEE, 2011.

[34] O. Woodman and R. Harle. Pedestrian localisation for indoor environments. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 114–123. ACM, 2008.