Indoor Navigation with Mixed Reality World-in-Miniature Views and Sparse Localization on Mobile Devices

Alessandro Mulloni, Hartmut Seichter, Dieter Schmalstieg

Institute for Computer Graphics and Vision, Graz University of Technology, Austria

[mulloni | seichter | schmalstieg] @ icg.tugraz.at

(a) (b) (c)

Figure 1. We support indoor navigation with a Mixed Reality World-in-Miniature (WIM) and sparse localization. As a user walks, we use a Virtual Reality WIM to illustrate the next instruction (a). Once the user reaches an info point (a floor-mounted poster) we show the whole path using an Augmented Reality WIM (b). After departing the info point, we resort again to Virtual Reality (c).

ABSTRACT

We present the design of an interface that provides continuous navigational support for indoor scenarios where localization is only available at sparse, discrete locations (info points). Our interface combines turn-by-turn instructions with a World-in-Miniature (WIM). In a previous study, we showed that using an Augmented Reality WIM at info points, and turn-by-turn instructions elsewhere, is a valid support for navigation inside an unknown building. In particular, we highlighted that users value the WIM as a tool for monitoring their location in the building. In this work, we focus on using the WIM continuously, not only at info points, to support navigation. We adapt the WIM views to the quality of localization by transitioning within Mixed Reality: we use Augmented Reality to provide an overview of the whole path at info points and Virtual Reality to communicate the next instruction when localization is not available. Our results from a new user study validate our interface design and show that users exploit not only turn-by-turn instructions but also the WIM throughout the path, to navigate with our interface. This paper provides insight on how a low-infrastructure indoor solution can support human navigational abilities effectively.

Categories and Subject Descriptors

H.5.1 [**Information interfaces and presentation**]: Multimedia Information Systems – *Artificial, augmented, and virtual realities.* H.5.2 [**Information interfaces and presentation**]: User Interfaces.

General Terms

Design, Experimentation.

AVI '12, May 21-25, 2012, Capri Island, Italy

Keywords

Augmented reality, mixed reality, indoor navigation.

1. INTRODUCTION

Many indoor navigation systems are based on the assumption of continuous localization of the user. Yet, this is not always viable, particularly indoors, as it requires complex instrumentation of the environment or expensive sensors. Localization only at few discrete points in the building (*sparse localization*) requires less instrumentation of the environment, but sparse localization systems typically provide detailed instructions only at selected *info points*, and no instructions elsewhere. One underlying assumption for continuous localization is that users may need prompt support during navigation anywhere, not just at info points. We argue that human navigation abilities can be exploited when continuous localization is not viable: humans can follow sequences of turn-by-turn instructions, if sufficient context information is provided.

In a previous work [11] we showed that Augmented Reality (AR) World-in-Miniature (WIM) views at info points, combined with turn-by-turn instructions between info points, are a valid support for navigation in an unknown building. We also observed how users monitor their current location by matching the interface view with the surrounding environment. Users valued the support given by the WIM to this matching task at the sparse info points, while they lamented the lack of such support elsewhere.

In this work, we present a refined interface design. Based on our previous findings, we emphasize the role of the WIM: in contrast to our previous design, we *continuously support users' navigation with WIM views*, not only at info points, by transitioning within Mixed Reality (MR) between Virtual Reality (VR) and AR (Figure 1). We present results from a new user study that show how users succeed in navigating indoors with our interface. Further, we verify that with the new interface design users exploit the WIM to navigate, not only at info points. Our results confirm that turn-by-turn instructions act as strong navigation cues, but also highlight that landmarks in the WIM act as prominent secondary cues.

2. RELATED WORK

A key requirement for indoor navigation systems is localizing the user within the building. Sensing infrastructure can be used, for example infrared beacons [1] or radio frequency (RFID tags) [5]. Wagner and Schmalstieg [12] use computer vision. These solutions require instrumentation of the environment with emitters, sensors or markers. An alternative approach is dead reckoning: incrementally updating the user's position by measuring her movements. Dead reckoning typically drifts, losing accuracy over time. Foxlin [6] for example achieves a drift of 0.3% of travelled distance using shoe-mounted sensors. Dead reckoning is often combined with other approaches: for example, Löchtefeld et al. [10] combine it with manual recalibration to reduce drift. Their approach is very flexible and works with arbitrary maps in unprepared buildings. Yet, it requires manual user intervention at initialization time (dead reckoning cannot give an absolute location in the building) and every once in a while (because of drift in the dead reckoning). An extensive coverage of indoor localization technologies is beyond the scope of this paper, and we refer the reader to the survey by Hightower and Borriello [7] for further information.

Interfaces inherently reflect the accuracy of the underlying localization technology. Interfaces based on dense localization can present continuous feedback to the user. For example, Chittaro and Nadalutti [5] use RFID to continuously match the position and orientation of the user in a building with a virtual viewpoint in a 3D model of the same building. Their results from a user study are encouraging, as all users were able to navigate a given path using the 3D model. If localization accuracy varies, the interface must reflect such variations. For example, Höllerer et al. [8] use AR arrows and a WIM and – similarly to our work – propose an adaptive interface that transitions between AR and WIM views, when the localization accuracy is respectively high or low. Yet, in contrast to our work, they do not discuss the case in which there is no localization at all. Butz et al. [4] also discuss the case of varying localization accuracy in indoor navigation. Similarly to our approach, they suggest adapting the information shown on the screen based on localization accuracy, so that user's navigational ability can make up for the uncertainty of the system, when localization is inaccurate or absent.

In our work, we suggest exploiting human navigation abilities to have continuous support for navigation in a sparse-localization context. Humans are used to follow sequences of turn-by-turn instructions, if sufficient information is provided. In a recent paper, Brush et al. [3] show a navigation system that supports users with only a static list of turn-by-turn instructions: the results of their evaluation show evidence that this is a viable solution. But the lack of localization makes their approach unsuitable in situations where users deviate from the predefined path or get lost.

In a previous paper [11], we presented an interface that combines turn-by-turn instructions with AR clues at sparse localization points. In a user study, we validated our design and we observed that, during navigation, users monitor their location and progress by matching the interface view with the surrounding environment; supporting this matching task is a key design goal, and our previous results show that a WIM and the iconic representations of the instructions are a valid support for it. In this work, we present a refined design of our interface that enforces the role of the WIM as a continuous support for this matching. We also present results from a novel user study, which validates our last design iteration and provides further insight on how users exploit it in indoor navigation.

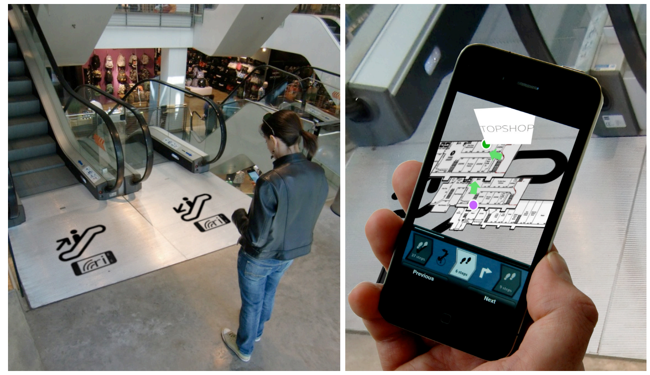


Figure 2. Conceptual design of info points for a shopping mall. Left: a user accessing the info point. Right: view on the phone.

3. INTERFACE DESIGN

To achieve the goal of continuously supporting users' navigation with WIM views, not only at info points, we propose smooth transitions within the MR space, depending on the availability of localization.

3.1 Info points

We use a low-key localization infrastructure by only placing sparse *info points* in the building. Info points provide robustness against path deviations or navigation errors: if users deviate from the pre-defined path or make an error, they can rejoin the correct path at the nearest info point. At each info point, we install a poster on the floor containing a pattern that can be detected and tracked using computer vision technology[1]. In the central part of the poster we also encode a unique identifier for the info point (a 9-bit BCH code). We also write the numerical ID of the info point prominently on each poster. The physical appearance of info points allows for a clear affordance, because they are prominent artificial reference points.

We stress that current info points are only experimental. Info points rely on a sufficiently textured floor that can be part of the overall building design. In Figure 2, we present a conceptual design that fits a modern shopping-center scenario.

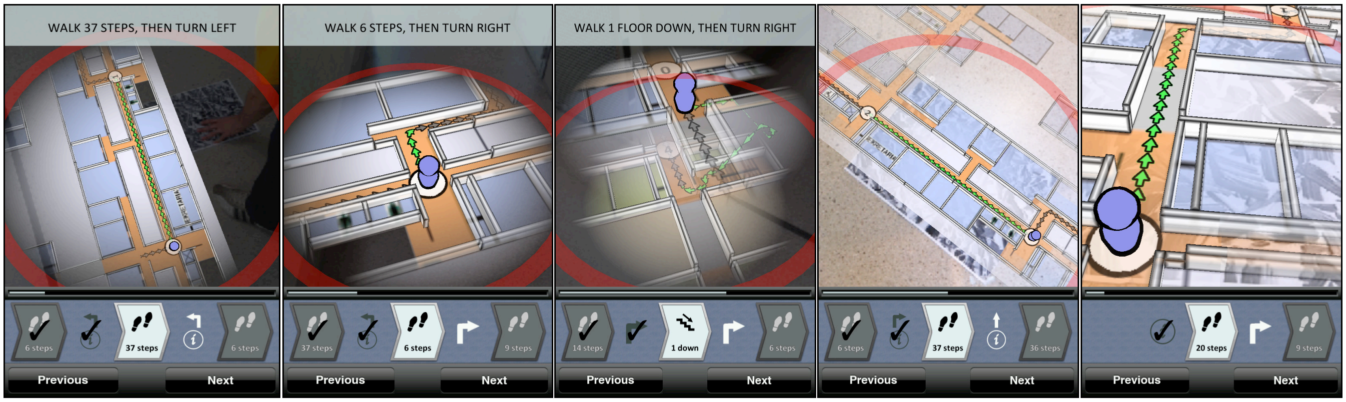
3.2 Interface view

Our interface splits the screen in two areas: the *instructions view*, which shows a list of turn-by-turn instructions, and the *MR view*, which illustrates the instructions using a Mixed-Reality WIM.

3.2.1 Instructions view

The instructions view (Figure 3, lower part of screenshots) shows a sequential list of what a user needs to perform in order to reach the target destination. The instructions supported are *walk*, *change floor*, *turn* and *reach office* and are shown as a sequence of arrow-shaped elements. We visualize turns and info points between instructions and we clearly identify with a checkmark which instructions have already been performed. Users can scroll through the instructions using the touch-screen. Users can also switch to the next or the previous instruction with a single tap – either on a button (on the bottom part of the instructions view) or on the instruction icon itself. On the upper part of the view, a small progress bar indicates the progress in the navigational task.

[1] We use Qualcomm's QCAR SDK as a solution for natural-feature tracking of the posters – http:// ar.qualcomm.com



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

Figure 3. World-in-Miniature views in our interface. To provide continuous navigation support with sparse localization, we transition within Mixed Reality: we use Virtual Reality when there is no localization (a-c) and Augmented Reality otherwise (d-e).

3.2.2 MR view

The MR view (Figure 3, upper part of the screenshots) details the instructions visually, using Mixed Reality WIM views. The WIM contains some of the landmarks of the building: walls, bridges, stairs, toilets and offices. Info points are also visualized. The WIM acts as a visual aid, so that users can match their view of the physical environment with the interface view on the screen. We adapt the WIM view depending on availability of localization.

If a user is walking between two info points, localization is not available. In this case we show a VR view, in which the WIM only details the current instruction (Figure 3 (a-c)), highlighting the next path segment to be traversed and the upcoming turn. If the user switches to the next instruction, a short animation moves the user's avatar (a blue pawn) through the path segment to the next intersection, and the view is then centered on the next instruction. The same applies if the user switches to the previous instruction. Textual information complements the WIM, providing verbal details on the current instruction. In the VR view, the viewpoint on the WIM is fixed: as we do not know the location of the user, we assume that she is on the path segment related to the current instruction and we maintain the focus on such segment. However, we use the phone's sensors to control the angle at which the WIM is viewed by tilting the phone. Parallax effects can help understanding the 3D structure of the WIM. For example, Figure 3 (a) shows a top-down view, while Figure 3 (b-c) show a more tilted view. In VR mode, video rendering is darkened rather than disabled to allow the user targeting an info point on approach.

When users approach an info point and they target it with the phone's camera, we transition from VR to an AR view that provides a more detailed overview of the whole path, highlighting the current position, the path to the destination, and the already traversed path (Figure 3 (d-e)). The visualization also shows the position of all info points within the building. We designed the WIM to appear as a 2D map from afar, but as a 3D model from closer, tilted points of view. The rationale is that a 2D map quickly conveys route information, whereas a WIM better supports landmark recognition as shown by Kray et al. [9]. A novel aspect of our approach is that we mount tracking targets on the floor, rather than on the walls: the AR WIM has therefore the same orientation of the physical building (no mental rotation is required by the users). Further, users can change the viewpoint on the WIM by simply moving the device over the tracking target, due to the AR interaction paradigm. In contrast, exploring the VR model in this detail would require a 6-degree-of-freedom manual control of the viewpoint on the WIM, which would be much more cumbersome.

Our interface runs interactively on an iPhone 4. In VR mode, gyroscope, magnetometer and accelerometer data are fused with linear Kalman filters to estimate the orientation of the device. Our implementation uses GLSL ES shaders for rendering.

4. USER STUDY AND DISCUSSION

We validated the combination of AR with turn-by-turn instructions in a previous experiment [11]. In this work, we conducted a new experiment focused on validating our refined interface design, and evaluating if the WIM effectively supports users in monitoring their position, not only at info points.

8 participants (4m/4f), aged between 24 and 30, took part in the experiment. Participants were asked to navigate a significantly complex path within our department's building (Figure 4) – a task requiring one floor change, crossing a skyway, 7 changes in heading and 9 wayfinding decisions at intersections.

We adopted a *think-aloud* approach: we asked participants to *state loud all their navigational decisions and why they took them*, during navigation. By choosing think-aloud, we aimed at collecting information on how users exploit environmental and interface cues during the task to make navigational decisions. We also tried to avoid the influence of post-task reasoning, which would have emerged more prominently with a post-task interview. All participants successfully completed the task.

At the end of the task, we collected usability data using the SUS (System Usability Scale) questionnaire [2]. The average score from SUS (on the range 0–100, 100 being the highest usability) is 75.31 (median = 83.75, $\sigma = 20.29$). The high standard deviation is inevitable, due to the small sample size and the subjective nature of the questionnaire. The result shows that participants did not have major usability issues with our system.

We transcribed all justifications of navigational decisions made by the participants and divided them into 10 groups, one for each path segment (see Table 1). We then extracted all the keywords related to navigation (e.g., "turn left"), landmarks (e.g., "door") or other spatial reasoning (e.g., "dead end"). Finally, for each path segment, we counted the utterances of each keyword. The results of this process are presented as a tag cloud in Figure 4. In Table 1, we distinguish between keywords that appear in the interface (top) and those that do not appear in the interface (bottom). Clearly, keywords cluster in proximity of corresponding building elements.

As expected, a large part (81%) of the keywords used by participants appear in textual form in the turn-by-turn instructions. In our previous experiment, we observed that turns and step counts

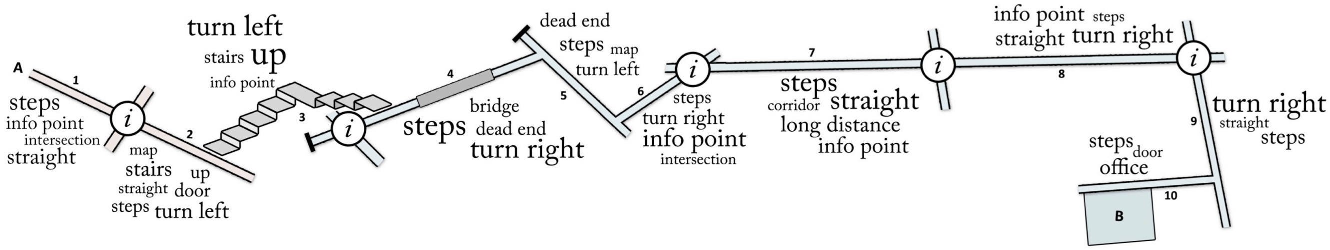


Figure 4. Users walked the path from A to B. We divide the path in 10 path segments with one floor change (3), one skyway (4) and various intersections. We asked participants to think aloud during navigation: keyword utterances are shown by the tag clouds.

are strong navigation indicators: our results from this experiment confirm this finding. Info points acted as strong landmarks: the posters (and the relative ID) appeared prominently on the floor of the building, in the WIM and in the instructions view. Some participants referred to them stating the specific number (e.g., "I am at info point 2"). The exact term "info point" did not appear in the interface, but the examiner used it in the introductory phase of the experiment. Nevertheless, one participant called it aptly a "check point", hinting at its role for confirmation and overview.

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

While keywords from turn-by-turn instructions are prominent in participants' justifications, the results suggest that there was also an underlying process, supported by the WIM, of matching the environment with the interface. Further, keyword utterances occurred throughout the path, not only at info points.

Many auxiliary landmarks (e.g., plants, fire extinguishers, tables) were missing in our WIM visualization: indeed, such landmarks also did not appear as keywords in participants' statements. We conclude that the level of detail of our WIM was sufficient for our specific navigation scenario: a compromise must be found between the modeling work necessary to produce the WIM, and the landmarks in the WIM needed by a user in a specific scenario.

5. CONCLUSION

Our evaluations confirm the validity of our design: combining MR WIM views with turn-by-turn instructions can support indoor navigation if continuous localization is not viable. Further, our results also confirm that providing informative views – in the form of a WIM – supports users in matching the interface with the physical environment. In line with our previous observations, and with work by Butz et al. [4], our results show that users' navigational ability can be exploited to substitute continuous localization, if a sufficiently informative and consistent interface is provided. An important aspect hereby is that transitioning within MR allows us to keep the WIM always on screen, adapting its visualization to the current availability of localization.

We concur from our validations that we reached a sufficiently effective design for combining Mixed Reality exocentric views to support indoor navigation. Yet, we suspect egocentric cues (first-person views of the WIM and augmentations of the physical environment) to be equally important. For the next iterations in our interface design, we therefore focus on possible combinations of our WIM views with egocentric MR views. Especially intriguing is how much an egocentric cue can contribute to the support of indoor navigation over existing modes. Looking ahead, this also suggests investigating in detail the underlying cognitive processes that occur using our interface, such as spatial learning and spatial reasoning, as they directly inform the interface design.

		Path segment										Total	
		1	2	3	4	5	6	7	8	9	10		
Keywords that appear in the interface	Steps	4	2		6	3	2	5	1	3	3	29	94
	Straight	3	1					5	3	1		13	
	Turn right				5		3		4	5		17	
	Turn left		3	5		2						10	
	Up		2	7								9	
	Office										3	3	
	Info point	2		1			4	3	3			13	
Keywords that do not appear in the interface	Intersection	1					1					2	22
	Stairs		3	2								5	
	Door		2								1	3	
	Bridge				2							2	
	Dead end				2	2						4	
	Corridor							1				1	
	Long distance							3				3	
	Map		1			1						2	

Table 1. Keyword utterances per path segment and in total.

6. ACKNOWLEDGMENTS

This work was supported through the Christian Doppler Laboratory for Handheld Augmented Reality.

7. REFERENCES

[1] Abowd, G.D. et al. 1997. Cyberguide: a mobile context-aware tour guide. *Wireless Networks*. 3, (Oct. 1997), 421–433.

[2] Brooke, J. 1996. SUS – A quick and dirty usability scale. *Usability evaluation in industry*, 189–194.

[3] Brush, A.J.B. et al. 2010. User experiences with activity-based navigation on mobile devices. *Proceedings of MobileHCI 2010*, 73–82.

[4] Butz, A. et al. 2001. A hybrid indoor navigation system. *Proceedings of IUI 2001*, 25–32.

[5] Chittaro, L. and Nadalutti, D. 2008. Presenting evacuation instructions on mobile devices by means of location-aware 3D virtual environments. *Proceedings of MobileHCI 2008*, 395–398.

[6] Foxlin, E. 2005. Pedestrian Tracking with Shoe-Mounted Inertial Sensors. *IEEE Computer Graphics and Applications*, vol.25, no.6, (Nov.-Dec. 2005), 38–46.

[7] Hightower, J. and Borriello, G. 2001. Location systems for ubiquitous computing. *Computer*. 34, 8 (Aug. 2001), 57–66.

[8] Höllerer, T. et al. 2001. Steps Toward Accommodating Variable Position Tracking Accuracy in a Mobile Augmented Reality System. *Proceedings of AIMS 2001*, 31–37.

[9] Kray, C. et al. 2003. Presenting route instructions on mobile devices. *Proceedings of IUI 2003* (Miami, Florida, USA, 2003), 117–126.

[10] Löchtefeld, M. et al. 2010. PINwI: pedestrian indoor navigation without infrastructure. *Proceedings of NordicCHI 2010*, 731–734.

[11] Mulloni, A. et al. 2011. Handheld Augmented Reality Indoor Navigation with Activity-Based Instructions. *Proceedings of MobileHCI 2011*, 211–220.

[12] Wagner, D. and Schmalstieg, D. 2003. First steps towards handheld augmented reality. *Proceedings of ISWC, 2003*, 127–135.