An Evaluation of Augmented Reality X-Ray Vision for Outdoor Navigation

Arindam Dey*
Magic Vision Lab
University of South Australia

Graeme Jarvis†
Magic Vision Lab
University of South Australia

Christian Sandor‡
Magic Vision Lab
University of South Australia

Ariawan Kusumo Wibowo§
Magic Vision Lab
University of South Australia

Ville-Veikko Mattila¶
Nokia Research Center
Nokia

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

Figure 1: We have ported our recent AR X-ray prototype [16] to a mobile phone (a) and conducted an evaluation of its effectiveness for outdoor navigation, comparing it against standard mobile navigation applications. Participants had to complete a 900 meter route (b). Our core finding is that AR X-ray required significantly less context switches than other conditions. Heatmap visualizations indicate the number of context switches on the path, averaged over all participants: AR X-ray (c), North-up map (d), and View-up map (e).

ABSTRACT

During the last decade, pedestrian navigation applications on mobile phones have become commonplace; most of them provide a birds-eye view of the environment. Recently, mobile Augmented Reality (AR) browsers have become popular, providing a complementary, egocentric view of where points of interest are located in the environment. As points of interest are often occluded by real-world objects, we previously developed a mobile AR X-ray system, which enables users to look through occluders.

We present an evaluation that compares it with two standard pedestrian navigation applications (North-up and View-up map). Participants had to walk a 900 meter route with three checkpoints along the path. Our main findings are based on the analysis of recorded videos. We could show that the number of context switches is significantly lowest in the AR X-ray condition.

We believe that this finding provides useful design constraints for any developer of mobile navigation applications.

Keywords: Augmented Reality, Evaluation, Visualization, Augmented Reality X-ray, Mobile Phone, Map, Navigation

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities H.1.2 [Models and Principles]: User/Machine Systems—Human factors

1 INTRODUCTION

During the last decade, pedestrian navigation applications on mobile phones have become commonplace; most of them provide a birds-eye view of the environment. There are many mobile applications providing navigation information along with landmarks to show points of interest such as Google maps and Nokia's Ovi maps. While these applications are widely used, they only provide an exocentric two dimensional view of the environment.

On the contrary, an environmental image being a combination of immediate sensation and past memory, is considered to be a strategic link in the process of way-finding and is used to interpret information and guide action [9]. This fact motivates us to provide a pictorial representation of the destination along with the immediate environmental image into a pedestrian navigation system which current map applications fail to provide.

Recently, mobile Augmented Reality (AR) browsers have become popular, providing a complementary, egocentric view of where points of interest are located in the environment. These applications commonly show points of interest on top of the real world, irrespective of their actual visibility. This causes several perceptual problems; most importantly, as occlusion is the most important depth cue [18], distances are hard to estimate. We have previously presented several AR see-through vision systems [2, 16, 17], which aim to improve the perception of occluded objects.

This paper first describes how we have ported our most recent AR X-ray system [16] to a mobile phone. Based on this platform, we have conducted an evaluation that compares it with two standard pedestrian navigation applications (North-up and View-up mobile map). Participants had to walk a 900 meter route with three check-

---

*e-mail: arindam.dey@unisa.edu.au
†e-mail: fanged@gmail.com
‡e-mail: chris.sandor@gmail.com
§e-mail: kusay003@mymail.unisa.edu.au
¶e-mail: ville-veikko.mattila@nokia.com

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

points along the path. We collected a large quantity of data during these trials: logged tracking data, completion time, and videos. Our main finding is based on the analysis of recorded videos. We were able to show that the number of context switches is significantly lowest in the AR X-ray condition. We believe that this finding provides useful design constraints for any developer of mobile navigation applications (see Section 5).

1.1 Related Work and Contribution

In this section, we first discuss related evaluations of mobile navigation applications. Second, we discuss related work on AR X-ray visualizations and their evaluation. Finally, we highlight our contribution based on the discussion of related work.

Since the first mobile pedestrian navigation application [1] was presented around 15 years ago, many evaluations have been conducted in this space. Typical topics of these studies were comparing 2D to 3D maps and also introducing novel navigation cues. A 3D map was found to be advantageous over a 2D map [14, 6]. While, 3D maps provide a more realistic and volumetric representation of the real environment, 2D maps enhance the use of previous knowledge effectively and reduce cognitive load [11]; for example, for an expert 2D map user, these advantages are minimal [8]. Several studies investigated the enhancement of common navigation aids through tactile feedback: paper map [12] and mobile maps [13]. In the same spirit as us, Rukzio and colleagues have evaluated common navigation aids against new paradigms for navigation: public displays and a rotating compass [15].

Various AR prototypes were built to provide location-based information; for example for tourist guide applications. A core challenges in these browsers is to show occluded points of interest. We have previously implemented several AR X-ray prototypes to address this challenge by experimenting with different visualization techniques: edge-overlay [2] and saliency [16]. We have also experimented with space-distorting visualizations to remove occluder objects in an intuitive way [17].

While our AR X-ray systems aim to create photorealistic renderings of occluded points of interest, most other research has focused on symbolic representations. Livingston et. al [7] have evaluated such a system through depth perception tasks. We have also previously evaluated two see-through visualizations using a handheld display [4]. Recently, a zooming interface for AR browsers was evaluated with an orientation task [10]. However, we are not aware of any evaluations that have evaluated the effectiveness of AR X-ray as a navigation aid.

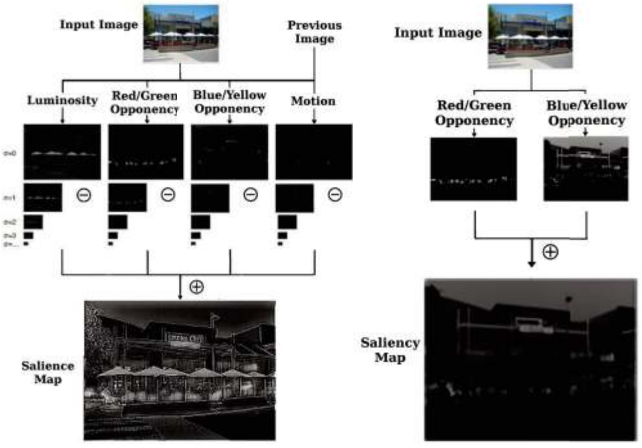
Contribution The core contribution of this paper is to present the first evaluation of an AR X-ray system in a navigation task. We could show that the number of context switches is significantly lower than with standard map applications on a mobile phone.

A side contribution of this paper is the porting of our previous AR X-ray system [16] to a mobile phone. This required us to perform several optimizations and adaption of our algorithms. Despite the limited computation power of mobile phones, we were able to achieve visually quite similar results to our previous prototype that ran on a laptop.

2 EXPERIMENTAL PLATFORM

We have ported our previous AR X-ray system [16] to the Nokia N900 mobile phone. Our original system ran on a laptop and was developed using Python and OpenGL 2.1. The N900 port uses C++ and OpenGL ES 2. Built on Qt4.7, the application runs as a plug-in module for Nokia's proprietary Mixed Reality Framework (MRF). MRF exposes the various sensors available on the N900 in a manner that is much easier to use than the native API, streamlining the initial setup of an AR application. The device's pose is determined by an externally attached sensor box connected via bluetooth. The sensor box provides an 'orientation' software sensor, a fusion of data from hardware sensors; compass and accelerometer.

Figure 2: Comparison of our previous prototype running on a laptop (left) and our N900 port (right). The visual appearance is quite similar.



(a) Laptop (b) Phone

Figure 3: Saliency computation

The porting of the AR X-ray system was successful, and produced results that are quite similar to our previous system running on a laptop (see Figure 2). In order to achieve an acceptable frame rate on the mobile phone, we had to perform three simplifications to our algorithm (see Figure 3); we removed three computations: mipmapping, motion saliency, and intensity saliency.

Mipmaps form an image pyramid, which provides multi-resolution images for feature detection and saliency calculation. This is a core part of the saliency calculation in our AR X-ray system, and must be run every frame. Benchmarking showed that mipmapping on our mobile phone accounted for approximately 500ms of rendering time per frame, which is significantly too slow. The reason for this (also confirmed by Nokia's driver developers) is that the mipmapping routine in the N900 is not very optimized. Typical mobile 3D applications, such as games, run the mipmapping step only once at startup, when loading the textures. We attempted several fixes to alleviate the mipmapping limitation without success including manual generation of mipmaps. Finally, as we could not overcome the performance problems, the mipmapping was removed completely, with the results being very comparable to using mipmapping previously.

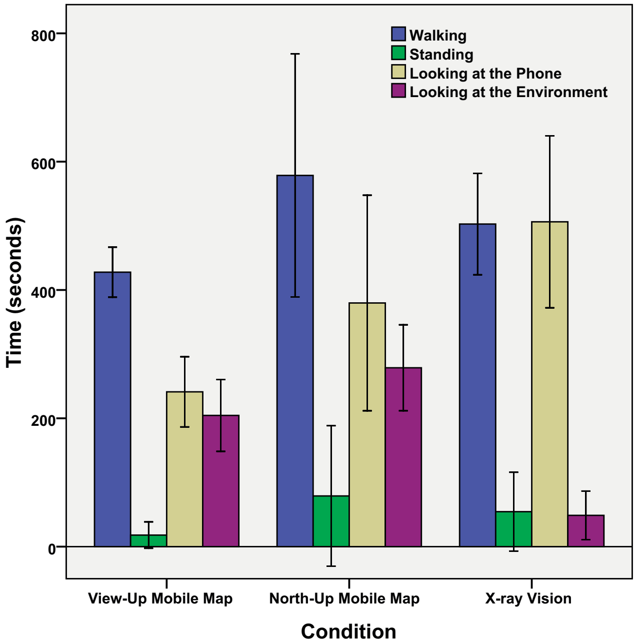


Figure 4: Video analysis: comparison of time spent in different activities. Compared to other conditions, participants looked significantly less at the environment and significantly more on the phone in the AR X-ray condition. Whiskers represent ±95% confidence intervals.

3 EVALUATION

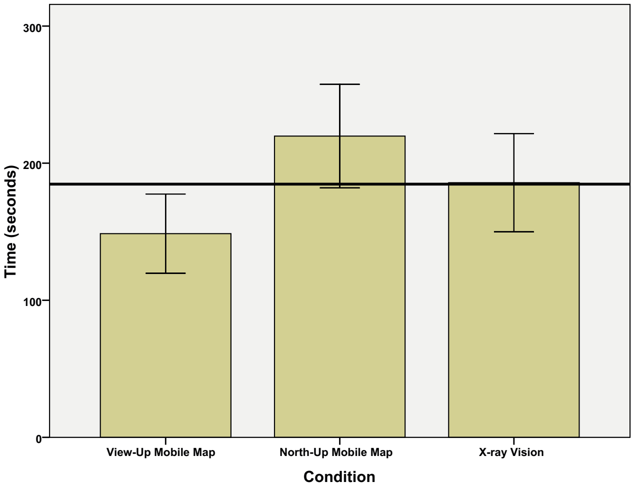
12 voluntary participants (all male) with ages ranging from 22 to 40 years were recruited from the student population of the University of South Australia. In a between-subjects design, we divided the 12 participants into three groups of four participants; each group was exposed to one of the three conditions, as described in Section 3.1. We selected three different target locations on the campus of the University of South Australia. The locations were carefully chosen to be among the least accessed in the campus. The average length of path segments was 289 meters (SD=91.8). Each participant traveled to all of the target locations in the same order using the assigned condition. The entire experiment took about 30 minutes per participant.

We instructed participants to navigate to the target location as they would have done normally in their day to day life. We did not specify any predefined path, as we wanted to investigate the difference in choice of paths using the different conditions. We asked participants to speak out loud while navigating.

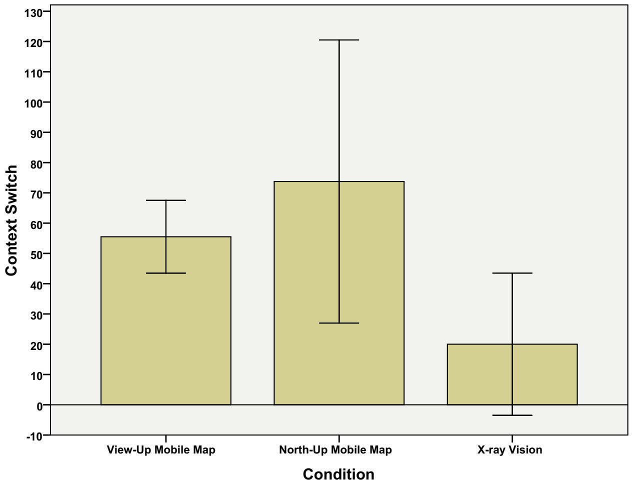
3.1 Conditions

AR X-ray vision as a navigation aid was the focus of our evaluation. In this condition, participants were provided with a mobile phone where only a photorealistic view of the target location was displayed through our AR X-ray vision. No other information such as route direction or distance to the target were provided. After participants reached a target location successfully, the next target location was loaded by the experimenter and presented to the participant. Overall, there were three target locations

As baseline conditions, we used two standard pedestrian navigation applications (North-up and View-up map). As North-up map, we ran Nokia's Ovi Maps on the N900. As View-up map, we used Apple's Maps application, which is preinstalled on iPhones, on an iPhone 3GS in View-up mode. We could not use Ovi Maps for this condition, as Ovi Maps does not support View-up maps. However, the appearance of both of the mobile maps were verified to have similar legibility. In the case of both of these mobile maps, only a target location was marked at one time on the map with a pin. Similar to the AR X-ray vision condition, once participants reached the location the next location was marked with a pin.



(a) Task completion time: Though there were no significant difference between AR X-ray vision and two other conditions; View-up map was significantly faster than North-up map.



(b) Video analysis: number of context switches. AR X-ray caused significantly less context switches compared to other conditions.

Figure 5: Further results. Whiskers represent ±95% confidence intervals and the thick Black lines represent overall mean.

We collected three different types of data: task completion time, GPS tracks, and video recordings. Task completion time was measured using a stopwatch. Our main data source were video recordings of participants. We externally recorded participants throughout their travel to the target locations using a Canon 550D camera at 60 fps. Later, we analyzed the video by identifying different behaviors of participants.

4 RESULTS

With regards to task completion time (see Figure 5(a)), the only significant difference was that View-up map was faster than North-up map (determined by a one-way ANOVA with $F(2,9) = 6.15; p = .017, \eta^2 = .58$). In the following, we focus on the results from the video analysis. We collected 243 minutes of video data for our 12

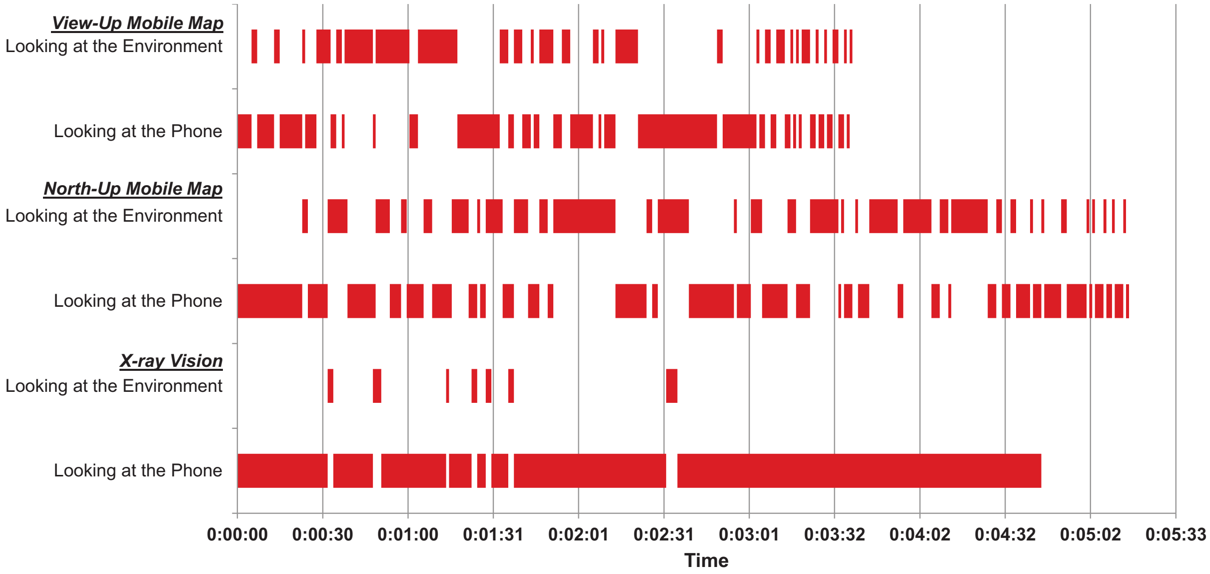


Figure 6: Video analysis: raw segmentation results for three typical trials. The number of context switches is clearly less for AR X-ray vision.

participants.

The video analysis yielded non-significant differences for disorientation and walking/standing across conditions. Walking/standing refers to the ratio of time that a participant spent in each mode. There was no significant interaction effect between walking/standing and the condition (see Figure 4). We define disorientation as participants standing at a fixed position for more than five seconds. All together, there were 82 occasions (View-Up map: 17, North-Up map: 40, AR X-ray: 25) when participants stopped while performing the navigation task. Out of these stops, the number of times when participants stopped for more than five seconds was: View-up map: 6, North-up map: 24, and AR X-ray: 13.

However, we could identify two significant effects in the video analysis: target of user's gaze (environment or mobile phone) and context switches. With regards to the target of the user's gaze, there was a significant main effect $F(2,9) = 6.21; p = .02; {\eta_p}^2 = .6$; in all conditions the environment was looked at less than the mobile phone. There was a significant interaction effect between condition and gaze target $F(2,9) = 25.56; p < .001; {\eta_p}^2 = .85$. In the AR X-ray condition, the gaze ratio of phone to environment was significantly higher than in any other condition (see Figure 4). A context switch was measured when participants switched their gaze from the mobile phone's screen to the environment. An one-way ANOVA showed a significant main effect of conditions on context switch $F(2,9) = 7.87; p = .011; \eta^2 = .64$ (see Figures 5(b), 6, and Figure 1(c-e)). AR X-ray had least context switch among all of the conditions. A Tukey's HSD post-hoc test revealed that the difference was significant (p=.009) with the View-Up map, but not with the North-Up map (p=.07). The average time after which a context switch occurred was: View-up map: 8.36 seconds, North-up map: 9.7 seconds, and AR X-ray: 31.6 seconds.

5 DISCUSSION

In this paper, we have presented the first evaluation of an AR X-ray system in a navigation task. In order to perform this study, we have ported our previous AR X-ray system to a mobile phone.

The most important result of our evaluation is based on the analysis of recorded videos. In the AR X-ray condition, the number of context switches is significantly lowest; additionally, participants looked significantly more at the mobile phone than at their environment. Even with a lower number of participants in our experiment, the results showed a higher level of effect size. This result is not surprising, as the AR view on the mobile phone enables users to observe their environment and the navigation cues simultaneously as there is no need to look at the environment directly.

The number of context switches is closely linked to attention: more context switches consume more of the user's attention. The amount of free attention is positively correlated with perceptual, cognitive, and motor tasks. In our experiment, the AR X-ray condition required less eye-movements due to a better spatial relation between stimuli (AR X-ray depiction of the target) and responses (walking direction), also known as stimulus-response compatibility. A better stimulus-response compatibility is known to enable the user to perform more accurate actions [3].

So, we believe that our result is valuable, as it indicates the benefits of AR as a navigation aid, which consumes less attention of the user; therefore, resulting in a more efficient navigation. Any other task that requires the user's attention simultaneously to be on the environment and at the same time to be on some additional information about the environment can benefit from AR as well; for example, maintenance instructions while performing maintenance [5].

In the future, we want to further investigate the possibilities of using AR as a mainstream navigation aid particularly for pedestrians. Additionally, we plan to further improve the speed of our prototype. We also plan to compare AR X-ray against standard AR browsers and other mobile map applications. It will also be valuable to validate our study in a city center with more participants. It will be interesting to perform similar experiments in different social circumstances such as busy streets, unfamiliar location, and time critical situations and investigate the differences in results. We believe that our photorealistic depiction of points of interest will aid users significantly to build a richer mental model of their environment. However, we would like to develop an optimized visualization for the mobile phones despite its low resolution and limited processing power.

ACKNOWLEDGEMENTS

The authors wish to thank the voluntary participants of the evaluation. This research was funded by Nokia Research Center.

REFERENCES

[1] G. D. Abowd, C. G. Atkeson, J. Hong, S. Long, R. Kooper, M. Pinkerton, and U. Centre. Cyberguide: a mobile context-aware tour guide. *ACM Wireless Networks*, 3:421–433, 1997.

[2] B. Avery, C. Sandor, and B. Thomas. Improving spatial perception for augmented reality x-ray vision. In *Proceedings of the IEEE Virtual Reality Conference*, pages 79–82. IEEE, 2009.

[3] R. Chua, D. J. Weeks, and D. Goodman. The human-computer interaction handbook. chapter Perceptual-motor interaction: some implications for human-computer interaction, pages 23–34. L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 2003.

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

[5] S. Henderson and S. Feiner. Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. In *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*, pages 135 –144, oct. 2009.

[6] K. Laakso, O. Gjesdal, and J. R. Sulebak. Tourist information and navigation support by using 3d maps displayed on mobile devices. In *Workshop on Mobile Guides, Mobile HCI 2003 Symposium*, 2003.

[7] M. A. Livingston, J. Swan, J. L. Gabbard, T. H. Höllerer, D. Hix, S. J. Julier, Y. Baillot, and D. Brown. Resolving multiple occluded layers in augmented reality. In *ISMAR '03: Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, page 56, Washington, DC, USA, 2003. IEEE Computer Society.

[8] R. Looije, G. M. te Brake, and M. A. Neerincx. Usability engineering for mobile maps. In *Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology*, Mobility '07, pages 532–539, New York, NY, USA, 2007. ACM.

[9] K. Lynch. *The Image Of the City*. The MIT Press, 1960.

[10] A. Mulloni, A. Dünser, and D. Schmalstieg. Zooming interfaces for augmented reality browsers. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, MobileHCI '10, pages 161–170, New York, NY, USA, 2010. ACM.

[11] A. Oulasvirta, S. Estlander, and A. Nurminen. Embodied interaction with a 3d versus 2d mobile map. *Personal Ubiquitous Comput.*, 13:303–320, May 2009.

[12] M. Pielot, N. Henze, and S. Boll. Supporting map-based wayfinding with tactile cues. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '09, pages 23:1–23:10, New York, NY, USA, 2009. ACM.

[13] M. Pielot, B. Poppinga, and S. Boll. Pocketnavigator: vibro-tactile waypoint navigation for everyday mobile devices. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, MobileHCI '10, pages 423–426, New York, NY, USA, 2010. ACM.

[14] I. Rakkolainen, J. Timmerheid, and T. Vainio. A 3d city info for mobile users. *Computers & Graphics*, 25:619–625, 2000.

[15] E. Rukzio, M. Müller, and R. Hardy. Design, implementation and evaluation of a novel public display for pedestrian navigation: the rotating compass. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, pages 113–122, New York, NY, USA, April 2009. ACM.

[16] C. Sandor, A. Cunningham, A. Dey, and V. Mattila. An augmented reality x-ray system based on visual saliency. In *Proceedings of the IEEE International Symposium of Mixed and Augmented Reality*, pages 27–36, Seoul, Korea, 2010. IEEE.

[17] C. Sandor, A. Cunningham, U. Eck, D. Urquhart, G. Jarvis, A. Dey, S. Barbier, M. Marner, and S. Rhee. Egocentric space distortion visualizations for rapid environment exploration in mobile mixed reality. In *Proceedings of the IEEE Virtual Reality Conference 2010*, pages 47–50, Waltham, MA, USA, 2009. IEEE.

[18] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004.