Zooming Interfaces for Augmented Reality Browsers

Mulloni Alessandro[1], Andreas Dünser[2], Dieter Schmalstieg[1]

[1]Institute for Computer Graphics and Vision
Graz University of Technology
Inffeldgasse 16, 8010 Graz, Austria
[mulloni I schmalstieg]@icg.tugraz.at

[2]The HIT Lab NZ
University of Canterbury
Private Bag 4800, Christchurch, New Zealand
andreas.duenser@hitlabnz.org

ABSTRACT
Augmented Reality combines real world and virtual information in interactive visualizations. Since phones started integrating GPS, compass and accelerometer, several Augmented Reality browsers for phones have hit the market. These are applications that access large amounts of geo-referenced information from online sources and present it at corresponding physical locations, superimposed onto a live video stream. However, Augmented Reality is constrained by the camera's field of view and restricted to first-person views, limiting the amount of overview that users can gain. We present two zooming interfaces that compensate for these constraints by enabling users to smoothly zoom between the Augmented Reality view and (1) an egocentric panoramic view of 360°, and (2) an exocentric top-down view. We present the results from two studies that show how in most search tasks our zooming interfaces are faster and require less panning than an overlay-based tool, scaling better as the amount of information grows.

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, Human Factors.

Keywords
Mobile Augmented Reality, Zooming Interfaces.

1. INTRODUCTION
Augmented Reality (AR) interactively combines digital information with the real world, typically superimposing the virtual information onto a specific spatial location in the live video stream. As modern mobile phones begin to integrate a wide range of sensors – GPS, digital compass, and accelerometer – they are becoming a viable platform for location-based AR applications. During the last year the first simple *"AR browsers"*, e.g. Wikitude[1] or Layar[2], hit the market. These are location-based systems that combine live video with virtual content retrieved from online sources on mobile phones, allowing users to visualize and interact with spatial information directly in its corresponding real-world location.

However, AR is bound to the field of view of the camera (typically 60° on mobile phones) and restricted to first-person views, whereas spatial information is spread in 360° around a user. Thus information might be outside the field of view and not visible in the AR visualization, limiting the amount of overview users can get. In general, three main approaches target the problem of presenting the information in context:

- *Overview+Detail* separates focus and context spatially, providing two distinct views.
- *Focus+Context* merges both the focus and the context into a single visualization, optimizing the use of the screen via distortions.
- *Zooming* separates focus and context temporally, allowing users to zoom and pan the visualization.

On mobile phones screen space is particularly scarce, so that temporal separation of views is often employed and especially important for visualizing large information spaces. We adopt this approach with two types of zooming interfaces for AR designed to compensate for the field of view and the first-person constraints (Figure 1). We propose (1) an egocentric zooming interface that enlarges the field of view up to a 360° panoramic image and (2) an exocentric zooming interface that transitions the user to a third-person view over the information. Both techniques run in real-time on an off-the-shelf smartphone.

2. RELATED WORK
The value in the combination of AR with geo-referenced content had already been envisioned in the late '90s. Egenhofer [8] was envisioning the *Spatial Information Appliances* as a family of applications that combine virtual reality, GPS, internet and mobile phones. In particular, Egenhofer presented the *Smart Glasses* as a tool to display information superimposed onto the real world. In the same year, Spohrer [19] described a similar concept, the *Worldboard glasses*, to overlay and align real and virtual objects.

Pioneering work by Rekimoto and Nagao, the *NaviCam* [16], was the first working, although tethered, handheld AR browsing device superimposing real-world objects with situation-sensitive information through the use of color-coded fiducials. Around the same time the *Touring Machine* [9] was the first real mobile AR system (using a backpack setup). The *Real-World Wide Web*

[1] http://www.wikitude.org/

[2] http://layar.com/

MobileHCI 2010 September 7-10, 2010, Lisboa, Portugal.
ACM 978-1-60558-835-3/10/09.

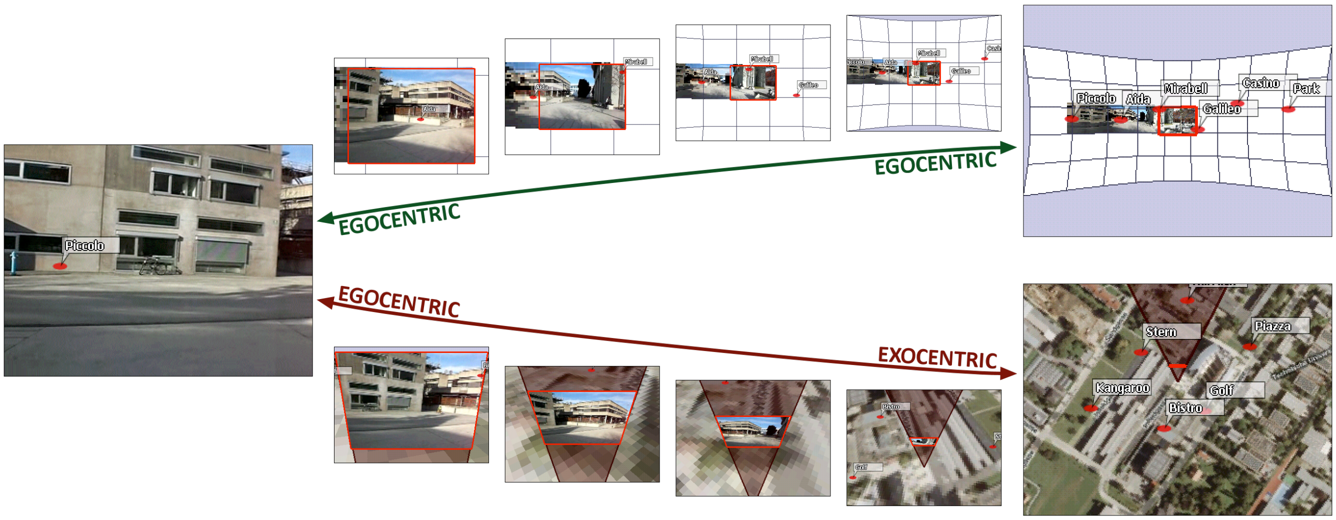


Figure 1. The proposed zooming interfaces, as the user turns the camera to the right. We propose an egocentric zoom that increases the field of view up to 360° (top) and an exocentric zoom that gives the user a top-down view onto the information (bottom).

Browser [14] was the first prototype of an AR browser, also based on a backpack setup, presenting information about the university campus by overlaying the live video with text labels. Since then a large number of mobile AR systems for various applications have been built, most recently also on smartphones [20]. Due to the availability of position and orientation sensors on modern mobile phones, during the last year there has been an increasing commercial interest in *"AR browsers"*, systems that interactively combine real video streams with virtual content retrieved from online databases, on off-the-shelf mobile phones.

Presenting information in context is one important objective of information visualization, and Cockburn et al. [7] present an extensive survey of possible approaches for context presentation. For AR browsers, Overview+Detail is the dominating approach. For example, Wikitude and Layar use radar-shaped glyphs to illustrate the position of information in the surroundings, and provide also an on-demand map view. Yet, Fröhlich et al. [10] provide preliminary evidence of users' dislike for a radar-like visualization in comparison with maps or AR. Lehikoinen and Suomela presented the *Context Compass* [15], a one-dimensional overlay that shows the horizontal angle of geo-referenced information with respect to the user's view direction.

Generally, results presented by Cockburn et al. [7] show evidence of cognitive effort when users are forced to switch between two separate views (Overview+Detail), as well as when users must assimilate state changes between pre-zoom and post-zoom conditions (Zooming). Yet, in the specific case of mobile devices, Chittaro [6] points out the difficulties of creating informative Overview+Detail visualizations due to the limited screen space; he suggests as possible alternatives hinting at off-screen objects, or employing intuitive navigation techniques.

Biocca et al. [3] propose the *attention funnel*, an AR visualization that uses a tunnel metaphor to guide the user's attention to a specific object, even if it is not currently visible. Schwerdtfeger and Klinker [18] show an application of the attention funnel for picking items guided by AR. Yet, the goal of these works is to pilot users' attention to specific objects and they are not suitable for providing an overview of the surrounding context.

Zooming interfaces allow smooth movement between overview and detail views. They have been often applied for visualizing and browsing large documents on small screens, where screen estate is particularly precious, for example when displaying scatter-plots [5], maps and web pages [4] on mobile devices. Zooming interfaces on phones have also been widely used commercially, for example for browsing web pages (e.g., Opera Mini[3]) or other general documents (e.g., Picsel's File Viewer[4]) on mobile phones.

In a recent work, Fröhlich et al. [11] show preliminary results on user preference for exocentric perspectives and larger fields of view when browsing geo-referenced information with a mobile device. Interfaces that allow zooming between egocentric and exocentric views have already been explored in AR. Grasset et al. [12] formalize the concept of a *transitional interface* as an interface for moving between representations and contexts. Their work mostly refers to the MagicBook by Billinghurst et al. [2], where transitions happen between an AR tabletop view and a VR immersive view. The zooming interfaces presented in this work also fall under the definition of transitional interfaces since they allow smoothly moving between AR visualizations and other visualizations, and between different points of view. An approach to exocentric zooming in AR similar to the one presented in this paper is suggested by Avery et al. [1]; yet, their work is mostly focused on X-Ray visualization and the usability of the zooming interface is not investigated.

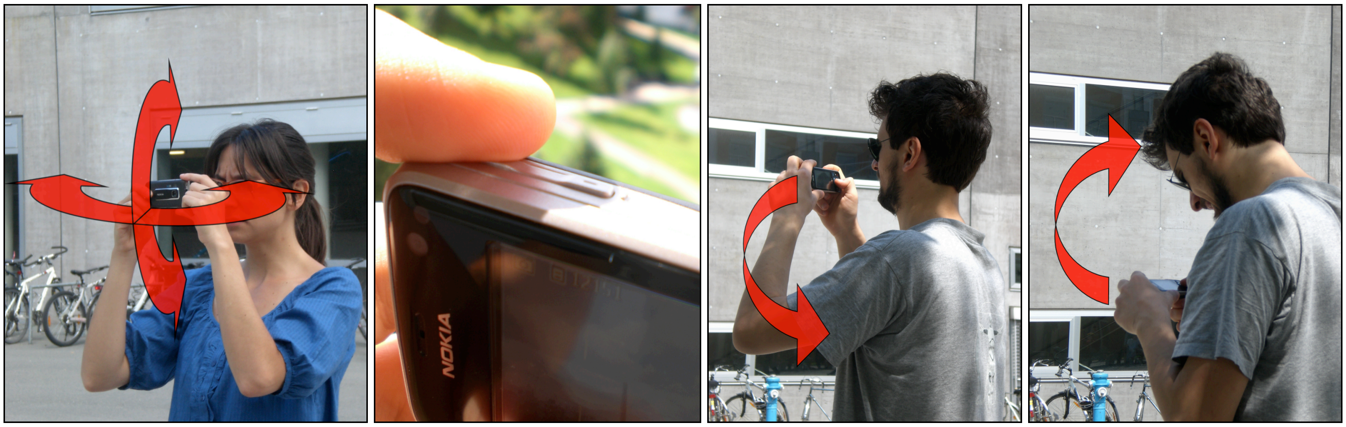
3. ZOOMING INTERFACES FOR AR

In an AR browser users will typically search for *points of interest* (POIs), e.g. a café given its name, or the nearest pharmacy. Users might also want to infer spatial relationships between POIs, e.g. the path from the hotel to the restaurant. Users will generally operate within two different reference frames – *egocentric* and *exocentric* – to build mental representations of the surrounding information.

---

[3] http://www.opera.com/mini/demo/

[4] http://www.picsel.com/flash/uploads/index.html



(A) (B) (C) (D)

Figure 2. Panning and zooming in the proposed zooming interfaces. (A) Panning is performed by physically turning the camera phone. (B) Zooming Panorama is triggered by the zoom buttons of the camera phone. Zooming Map is triggered by tilting the phone (C) down for the exocentric view or (D) up for the egocentric view.

Since AR is bound to the egocentric video feed, a naïve AR approach requires users to physically turn the camera in space fetching information and to mentally maintain a map of the information that has been visited. Interfaces for AR browsers should target these issues to improve their usability and effectiveness. We propose two types of zooming interfaces that provide the users with an egocentric zoom that increases the field of view up to 360° and an exocentric zoom that smoothly moves between first- and third-person views of the information.

Our proposed zooming interfaces replace the spatial separation typical of Overview+Detail with a time-separation approach. Users can interactively pan and zoom the visualization. Animations support coherence between zoomed-in and zoomed-out views, as suggested by Robertson: *"interactive animation is used to shift some of the user's cognitive load to the human perceptual system"* ([17], p.190).

The visualization of the *Zooming Panorama* is always centered on the live video of the phone. As users zoom out, gradually less screen space is used for the live video and more space is left for the surrounding context (Figure 1, top). POIs are shown on the screen as text labels indicating the identity of the object. As the camera rotates, we generate a panorama in real-time, using natural feature tracking in the video stream. More information on the tracking technology can be found in [20]. We visualize this panorama as a spatial clue for the user's rotations and as a trace of the visited information. The distortion of the panorama is a function of the horizontal and vertical angles relative to the user's view direction. We integrate a wireframe grid to support the understanding of the distortion as suggested by Zanella et al. [21].

The visualization of the *Zooming Map* is always centered on the user's position. As users zoom out, we smoothly animate the camera from the first-person AR view to an exocentric view presenting a satellite image of the user's surroundings (Figure 1, bottom). In the zoomed-out visualization we highlight the user's position and field of view as a glyph. The satellite image is augmented with text labels corresponding to the POI's identity. We use a forward-up map, i.e., the user's view direction always corresponds to the top of the screen.

AR users can pan the information by physically turning the camera in the environment. Zooming the panorama is triggered via the zoom buttons of the phone, similarly to a regular digital camera. Tilting the phone down zooms to the exocentric map view while tilting it up zooms back to full-screen AR (Figure 2).

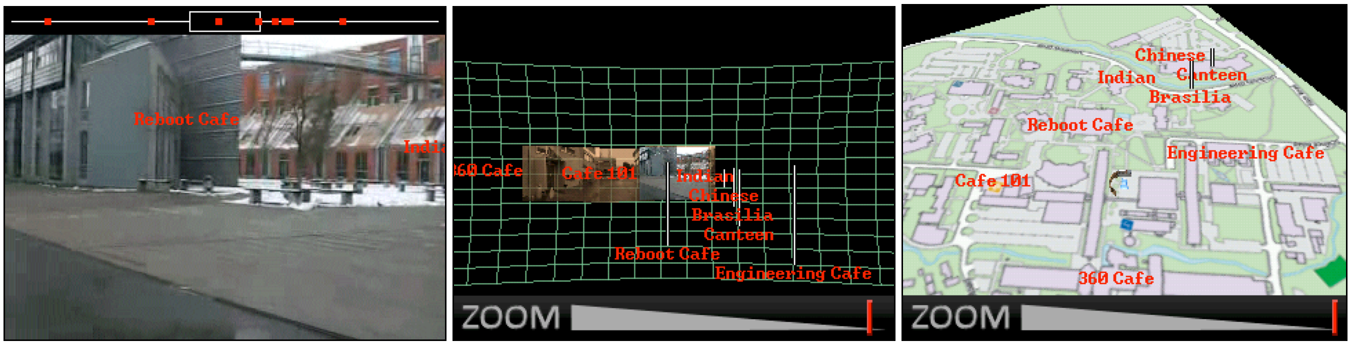
We implemented a prototype of our zooming interfaces using Studierstube ES, a software framework for AR applications on mobile phones, and interleaved the development with a number of expert evaluations in order to refine the interface. We also implemented the Context Compass [15]. In our implementation an overlay is used to show the horizontal position of POIs with respect to the user's view direction around 360°.

4. EXPERT EVALUATION

We implemented an early prototype of the interfaces to gather preliminary feedback on their usability. The prototype was running at interactive frame rates (approximately 10 frames per second) on an HP iPAQ 614c phone. Due to the lack of an accelerometer and a compass in the mobile phone, the tracking technology relied solely on optical flow from natural features. Figure 3 shows screenshots of the three interfaces in their early form.

Using this prototype, we conducted an exploratory study on the zooming interfaces at the main campus of our university. The study focused on identifying major usability problems of the proposed interfaces, on collecting subjective opinions, and on comparing the zooming interfaces with an alternative overlay-based approach (our custom implementation of the Context Compass). We recruited five expert users with experience in user interface design and augmented reality. The users had background in computer science (2), architecture (1), psychology (1) and arts (1). The reason for this choice was guided by the fact that we required users to actively help us spotting initial usability issues and to operate a non-robust tracking system.

Users were provided with a one-page information sheet presenting each interface with a screenshot, describing the visualization and the interaction. The prototype was programmed with geo-referenced positions of all cafés and restaurants on campus. After reading the information sheet, users were asked to spend some minutes with each interface separately, exploring the locations of



(A) (B) (C)

Figure 3. Early prototype used for gathering preliminary feedback: (A) Compass, (B) Zooming Panorama and (C) Zooming Map.

all cafés in the campus with respect to their location on the campus. We changed the order in which the interfaces were shown to different users. Our main interest was letting the users experience the potential uses and potential problems of each interface.

After trying the three interfaces, we conducted a semi-structured interview with each user, asking for opinions about pros and cons of each interface and pair-wise comparisons between the interfaces. During the interview, all users were provided with the information sheet containing the user interface screenshots, in order to support their memories.

4.1 Zooming Panorama

In the Zooming Panorama, the panorama trace left by the moving camera was generally found to be a useful cue. However, users commented that the panorama is "too dim" and merged with the black background, in particular with outdoor lighting conditions. Users were also concerned about the screen space occupied by the context in the zoomed-out view, claiming that it is hard to see the details in the video and in the panorama trace.

A recurring comment on the Zooming Panorama is that it encodes only direction information, while it does not provide information about distance and occlusions. Two users suggested considering adding representations for distance and occlusions (through color, transparency, etc.), commenting that a textual representation of the distance might clutter the view.

Users pointed out similarities between the Zooming Panorama and the Compass. Comparing the two, users noted both advantages and disadvantages in our proposed technique. While the Compass is limited in telling you where off-screen objects are, the Zooming Panorama also conveys what these objects are. In contrast, the Zooming Panorama takes more screen space and therefore the details can be hard to read when zooming out. Two users suggested combining the strengths of the two interfaces by showing the Compass overlay when the Zooming Panorama is completely zoomed in, in order to still provide some coarse context information. Comparing the Compass against the other two interfaces, one user pointed out that the compass has an advantage in that no user input is required.

4.2 Zooming Map

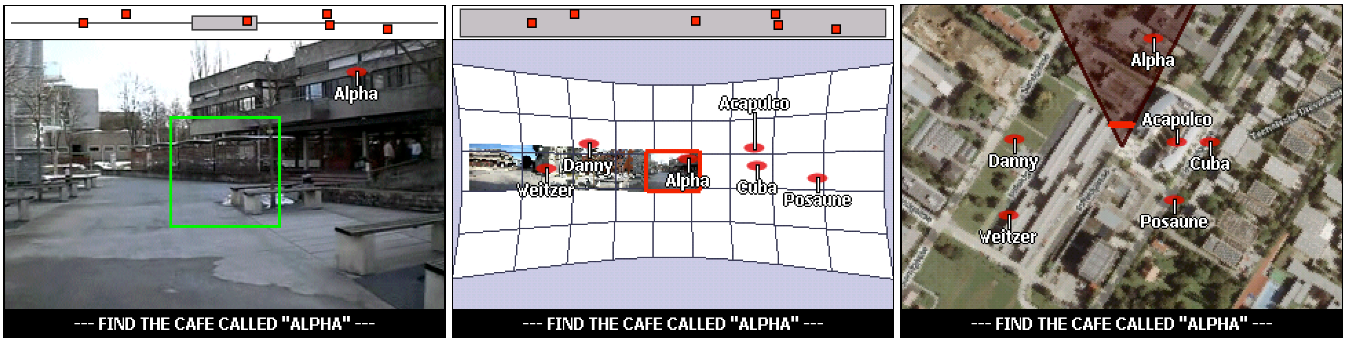
For the Zooming Map, most users stressed the importance for the transition to be smoother. One user suggested exploiting the smooth transition to "make the user aware that [the exocentric view] is a map, and it is flat or tilted down". A strong advantage of this interface seems to be its capability to show "the actual location, not just the orientation" of places, with words such as "distance" and "depth" spontaneously recurring in most interviews. Two users commented that the Zooming Map would be most useful when "you're stuck in a city […] not knowing what's around you" or "in a really difficult navigation space".

The Zooming Map was the most appreciated by all users. When comparing it with the other two interfaces, users generally claimed that their preference for it depends on the better level of overview provided by the map, in particular when a desired point of interest is not directly visible from the user's position. The ability to convey a measure of the distance from the user was also considered significant. Interestingly, the study revealed that most of the time users only exploited completely zoomed in and completely zoomed out views.

There were two contrasting opinions on the Zooming Map. One user said that she found the system easy to use because it fits the real-world experience that users already have with maps: In particular, this was put in contrast to the Zooming Panorama that adopts a non-familiar metaphor (a distorted panorama). A further advantage was seen in the fact that the map is user-centered and self-orienting (forward-up) making it quite different from a static map. A second user, however, was skeptical about the use of a map, saying that the technique might not be suitable for people that have problems with using real maps.

4.3 Refined prototype

Using the output of the preliminary study, we refined our prototype (Figure 4). We modified the Zooming Panorama in order to increase the contrast: The bright background, with darker grid lines, seems to improve the prominence of the panorama trace, also under outdoor-lighting conditions. The details in the panorama are still too small, but we didn't investigate possible solutions to the issue yet: Therefore, we exploit the panorama as a spatial trace rather than an augmentation target. As suggested by our users, we integrated the Zooming Panorama and the Compass into a single interface. We slightly modified the compass to also provide the vertical displacement of information, since we noticed that the previous 1D representation still required users to blindly fetch the information on the vertical axis. We changed the background to a bright color for the Compass too. Figure 4 (A-B) shows the new implementations, and how the Zooming Panorama and the Compass are combined and synchronized. The central rectangle that represents the current field of view of the user always matches the field of view of the Zooming Panorama. We



(A) (B) (C)

Figure 4. Prototype used for the user studies: (A) Compass, (B) Compass and Zooming Panorama and (C) Zooming Map. The user's task was always presented on the bottom of the screen. The green viewfinder in (A) represents the area for selecting the target object in order to complete a task.

didn't modify the Zooming Map but rather optimized it in order to speed up the rendering on phone, therefore making the animation smoother.

Based on the observation that with the Zooming Map users mostly used the completely zoomed-out or zoomed-in views, rather than intermediate viewpoints, we adopted the gesture-based approach already described in the previous section: tilting the phone down triggers a zoom to exocentric, while tilting it up triggers a zoom in to full-screen AR.

The refined prototype runs at approximately 15 frames per second on a Nokia 6210 Navigator phone. We use the phone's accelerometer, compass and GPS sensors to determine the user's orientation and position in space. We reduce the jittering of the sensors' output using a linear Kalman filter [13]. For slow camera movements we resort to optical flow, because it provides more accurate measurements compared to the phone's sensors. The phone's accelerometer is also used to detect the tilts that trigger the Zooming Map.

5. FIRST USER STUDY

Using the refined prototype, we conducted a study to investigate how our zooming interfaces perform in search tasks that we consider representative of a user on the move. We compared our interfaces against the Compass, designing the study on the following questions: Are the zooming interfaces faster? Do they require users to pan less in the information space? Do task-completion time and distance panned scale better as the amount of information increases?

5.1 Design

Twenty university students (10 female and 10 male) aged between 23 and 34 (*Mean* = 27.35, *Standard Deviation* = 3.10) participated in the study. All participants had normal or corrected-to-normal vision.

We designed four interfaces, each covering some of the three degrees of freedom of information surrounding a user – yaw, pitch and distance: Compass (yaw), Compass and Zooming Panorama (yaw, pitch), Compass and Zooming Map (yaw, distance), Compass, Zooming Panorama and Zooming Map (yaw, pitch, distance). We will refer to them as C, CP, CM and CPM. Every participant used all four interfaces, but the order followed a balanced Latin square to reduce carry-over effects.

In the study we asked users to perform two different search tasks: (1) find a well-defined café, with a name known a priori, and (2) find the closest café, therefore considering all information before making a choice. For each task we included two levels of information density (6 or 12 cafés).

For each of the four interfaces and the two tasks, the users performed one practice trial and two repetitions for two density levels. This resulted in 4 (Interface) x 2 (Task) x 2 (Density) x 2 (Repetition) = 32 observations per participant. A complete experimental session took approximately 45 minutes.

We represented cafés' locations as text labels containing the name and, for the second task, the distance from the user. To prevent learning effects and prior knowledge we randomized café names and locations, with distances varying between 40 and 160 meters.

Participants started each repetition by pressing the phone's center button, turned the phone in the direction of the correct café and pressed the center button again to complete the repetition (wrong presses were ignored). All users were instructed to work as fast and accurate as possible. We logged all task completion times and rotations of the camera (as angular distances), considering the latter as panning in the information space.

We formulated the following hypotheses: (H1) The zooming interfaces allow faster task completion times. (H2) The zooming interfaces require less panning compared to the Compass. We assume that users have to pan less since the zooming interfaces provide a more informative overview. (H3) No difference in completion time and panning between 6 and 12 cafés for the zooming interfaces will be evident, while the Compass will require more time and panning with 12 cafés than with 6 cafés.

5.2 Results

The collected data showed some outliers (5% for task completion time, 3.75% for panning). For all extreme outliers (3 x inter-quartile range), we kept the single non-outlying measurement of the two repetitions rather than their average. Figure 5 presents average completion time (top) and distance panned (bottom), for each combination of Interface, Task and Density.

For each task repetition, we divided the total distance panned by the shortest distance between the user's start orientation and the direction of the target bar. All distances are therefore presented as multiples of the shortest distance, where the ideal value is 1 (meaning that the user travelled the shortest possible path).

We analyzed the effects on time and distance with a 4 (Interface) x 2 (Task) x 2 (Density) repeated measures ANOVA.

Task completion time. All main effects were significant, as well as the interactions Interface x Task and Task x Density. Comparing the interfaces ($F_{2,\ 36.07} = 4.66$, $p = .02$) we found longer task completion times for C ($M = 27.84$, $SE = 2.58$) compared to CP ($M = 19.90$, $SE = 1.09$) while CM ($M = 24.21$, $SE = 1.73$) and CPM ($M = 22.12$, $SE = 1.71$) did not differ significantly from the other interfaces. Task 2 took longer on average ($M = 30.54$, $SE = 1.59$) than task 1 ($M = 16.45$, $SE = 0.61$) ($F_{1,\ 18} = 44.67$, $p < .01$) and high density ($M = 28.53$, $SE = 1.64$) took longer than low density ($M = 18.47$, $SE = 0.72$) ($F_{1,\ 18} = 40.78$, $p < .01$).

The interaction Interface x Task ($F_{3,\ 54} = 2.79$, $p = .05$) showed no difference between the interfaces for task 1, but for task 2 CP was faster than C and CM. For all interfaces task 2 took longer than task 1. Task x Density interaction ($F_{3,\ 18} = 17.33$, $p < .01$) showed a relatively steep increase in completion time for task 2 as the information density increases (+73%, 22.2 seconds for low density and 38.3 sec. for high density) compared to a smaller increase for task 1 (+27%, 14.5 sec. for low and 18.4 sec. for high density).

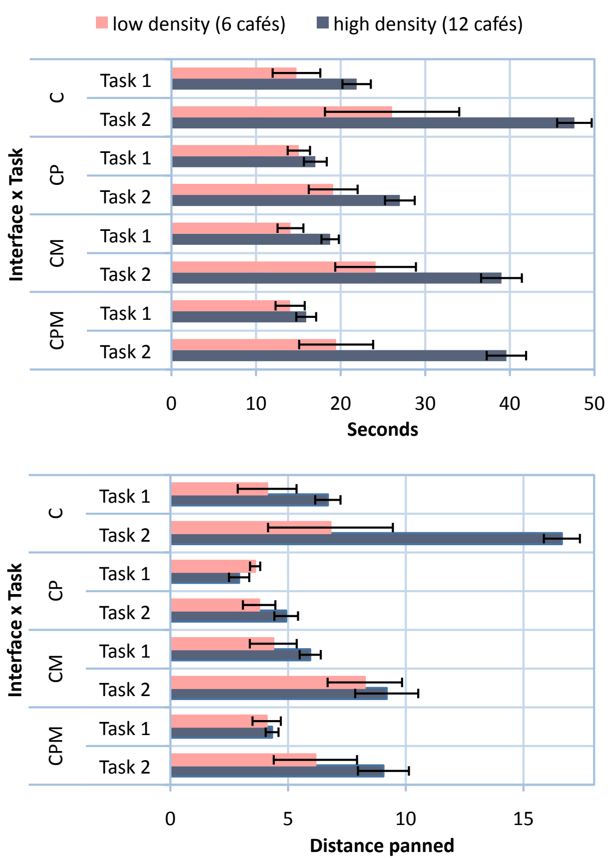


Figure 5. First user study: Task completion times in seconds and distance panned as multiple of the shortest distance (error bars = +/- SE), for each Interface, Task and Density.

Distance panned. All main effects were significant, as well as all two-way interactions. Comparing the interfaces ($F_{3,\ 54} = 10.62$, $p < .01$) we found that with CP the participants moved less ($M = 3.94$, $SE = 0.26$) than with C ($M = 8.86$, $SE = 0.95$), CM ($M = 7.40$, $SE = 0.66$) and CPM ($M = 5.97$, $SE = 0.59$). Task 2 showed almost double distance panned ($M = 8.50$, $SE = 0.60$) compared to task 1 ($M = 4.58$, $SE = 0.25$) ($F_{1,\ 18} = 46.18$, $p < .01$). Participants also rotated the phone more with 12 cafés ($M = 7.68$, $SE = 0.59$) than with 6 ($M = 5.41$, $SE = 0.33$) ($F_{1,\ 18} = 18.34$, $p < .01$).

The interaction Interface x Task ($F_{3,\ 54} = 2.77$, $p = .05$) showed no differences between the interfaces for task 1. For task 2, however, we found that CP required less panning than C and CM. CP was the only interface not showing a difference between the two tasks, while all other interfaces showed more panning for task 2 than for task 1. For the interaction Interface x Density ($F_{3,\ 54} = 5.86$, $p < .01$) we found less panning in the low-density case for CP compared to C and CM, while for high density CP was only different from C. The interaction Task x Density ($F_{3,\ 18} = 11.68$, $p < .01$) showed a bigger increase in panning distance between density levels for task 2 (+60%, 6.2 for low and 9.9 for high density) than for task 1 (+25%, 4.0 for low density and 5.0 for high density).

5.3 Questionnaire

After each interface the participants filled in a questionnaire, containing custom questions on ease of use, usefulness, and amount of information on the screen. Each statement was answered on a 6-point Likert scale ranging from “completely disagree” to “completely agree”. The questions and the results are shown in Figure 6.

The participants gave relatively high scores to the ease of use of all tested interfaces, for the low-density conditions. For the high-density conditions ratings were lower, with the CM interface getting the lowest rating. In terms of usefulness to complete the task, C received lower scores compared to the other interfaces in both the low and the high-density case. The two questions regarding the amount of information on the screen show more distinct patterns. For low-density tasks (6 cafés) the participants rated that there was neither too much nor too little information on the screen for all interfaces but C, where there was a slight trend towards too little information. For high density the Compass still shows slightly too little information, while the interfaces containing the Zooming Map tend to have slightly too much information. The users indicated that the animation between the views was generally helpful.

We also asked participants to rank the four interfaces from most preferred to least preferred. The Compass-Panorama-Map condition was ranked first, followed by Compass-Panorama, Compass-Map, and last the Compass ($\chi^2 = 35.22$, $df = 3$, $p < .01$).

6. SECOND USER STUDY

During the first user study we observed several instances in which the label-positioning algorithm failed, causing inconsistencies between subsequent frames. Some users commented on the issue at the end of the study. We speculate that most of the outlying points in the collected data are likely attributable to this problem. We fixed the labeling and conducted a study with smaller sample size to corroborate our previous findings with more robust data measurements.

We designed the study to also investigate on differences in completion time and panning between each separate interface. During the first study, we received informal feedback from users

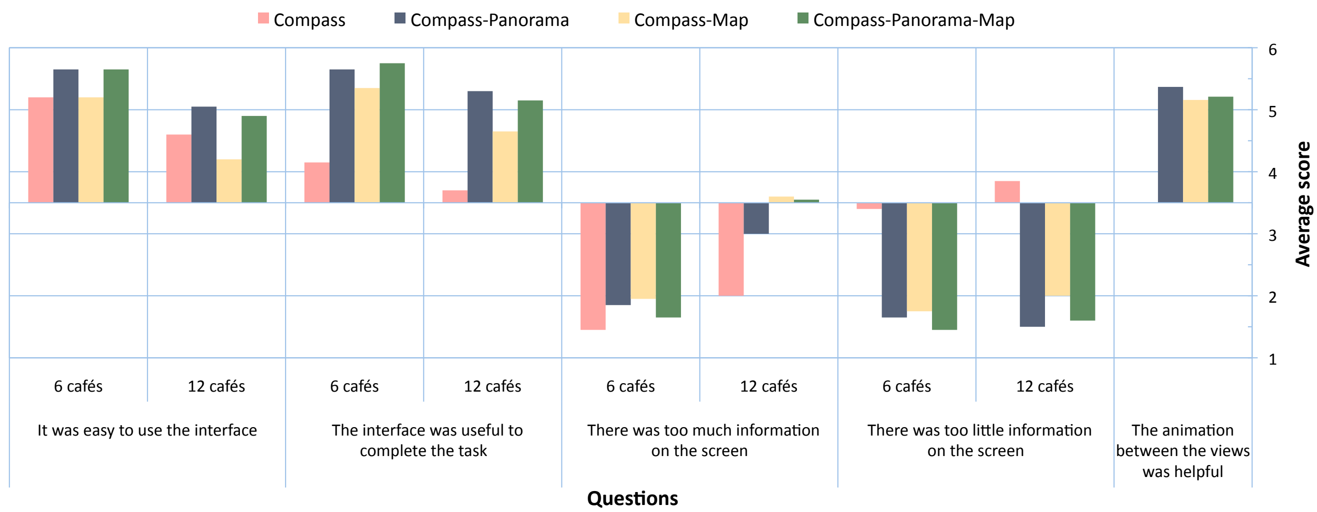


Figure 6. First user study: Questionnaire results. The subjective user ratings were given on a 6-point Likert scale, ranging from 1 (strongly disagree) to 6 (strongly agree).

on the possible benefits of having the Compass in the case of a single highlighted object. We therefore designed two search tasks: (1) find a highlighted object and (2) find the closest object. We formulated the following hypotheses: (H4) The zooming interfaces are slower than the Compass if the object is highlighted, but they require equal amounts of panning. (H5) The zooming interfaces are faster and require less panning than the Compass in task 2, when an overview on the data is needed.

6.1 Design

Ten people who participated in the previous user study were asked to join this study. We designed the interfaces as Compass (C), Zooming Panorama (P) and Zooming Map (M), to evaluate differences between each separate interface. We included two levels of information density, with either 6 of 12 labels. For task 2, the distance was again represented as a text label. The design of this evaluation followed that of the first user study.

6.2 Results

The effects of the experimental conditions on task completion time and distance panned were analyzed with 4 (Interface) x 2 (Task) x 2 (Density) repeated measures ANOVA. Distances panned were again normalized as multiples of the shortest distance to the target label. Figure 7 presents average completion time (top) and distance panned (bottom), for each combination of Task and Interface.

Task completion time. We found significant main effects for Interface and Task, and a significant interaction between the two. Task 2 took on average longer ($M = 15.81$, SE = 1.24) than task 1 ($M = 8.83$, SE = 0.49) ($F_{1,9} = 45.86$, $p < .01$). The main effect of Interface ($F_{2,18} = 8.57$, $p < .01$) showed that overall C ($M = 15.05$, $SE = 2.00$) took longer than M ($M = 10.81$, $SE = 0.67$). P did not differ from the other interfaces ($M = 11.11$, $SE = 0.50$). A closer look at the interaction Interface x Task ($F_{2,18} = 25.47$, $p < .01$) revealed that C was faster than the other two interfaces for task 1 and it was slower for task 2. Comparing the tasks over the interfaces shows a significant difference in completion time between tasks for C, but not for the two zooming interfaces.

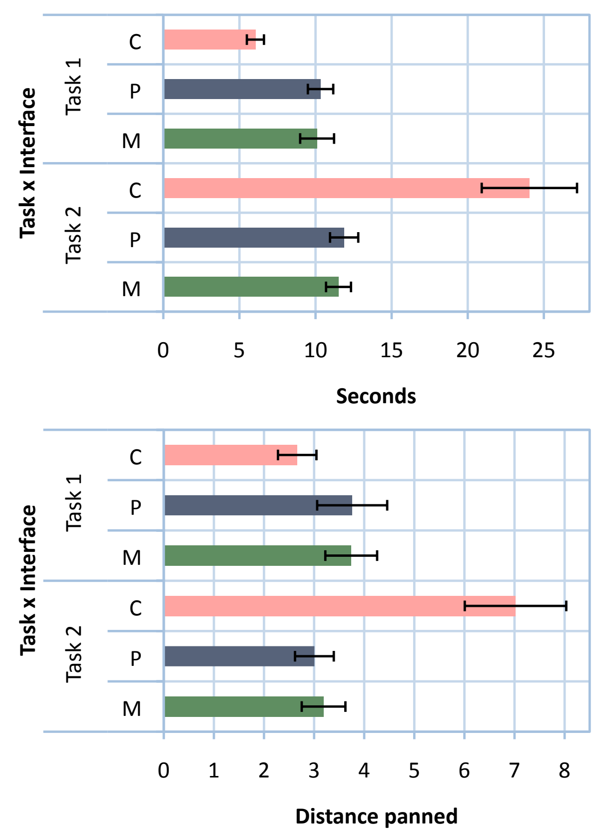


Figure 7. Second user study: Task completion times in seconds and distance panned as multiple of the shortest possible distance (error bars = +/- SE), for each Task and Interface.

Distance panned. We found a significant main effect for Interface ($F_{2,18} = 13.65$, $p < .01$) and a significant interaction between Interface and Task ($F_{2,18} = 19.37$, $p < .01$). Post-hoc analysis for the main effect showed overall more distance traveled for C compared to the other two interfaces. A closer look at the interaction Interface x Task showed that in task 1 the participants panned less with C than with M. For task 2, the distance panned with C was higher than with the other two interfaces. Comparing the tasks over the interfaces shows that there was only a difference between tasks for C, but not for the two zooming interfaces.

7. DISCUSSION

We conducted several studies to evaluate the usability and effectiveness of our proposed zooming interfaces. The expert evaluations aimed at uncovering usability issues, and they led to a refinement of the interfaces then used for two subsequent studies. The first study examined how users perform with the proposed zooming interfaces in typical search tasks and the second study backed up some of our previous results. These two studies compared the zooming interfaces with an overlay-based interface, the Compass.

We initially hypothesized that users would have been faster with the zooming interfaces than with the Compass (H1). Our study results only partially support this. In the first study we found the Compass to be slower than the combination of the Compass and Zooming Panorama interfaces, but users where not significantly faster when the Zooming Map interface was available. In the second study, the Compass took twice as long as the Zooming Map interface, and did not differ from the Zooming Panorama interface.

More revealing was the interaction between Interface and Task. We designed the tasks and interface conditions of the second study to measure how the type of task influences the performance with each interface. We hypothesized that the zooming interfaces would be slower than the Compass for a highlighted target object because of the time required for zooming, but we did not expect differences in distance panned (H4). For the second task we expected the zooming interfaces to be faster and to require less panning than the Compass, since they provide a more informative overview (H5). Both hypotheses were supported in our study, apart from the Zooming Map that required more panning than the Compass, for a highlighted object. It should be noted, however, that the Zooming Map interface requires the users to tilt the phone down to see the map, an additional movement adding to the panned distance.

The Compass was faster than the other interfaces with highlighted objects, but slower in the more complex search task. With the Compass participants took almost four times longer for task 2 than for task 1, whereas the increase in time between the two tasks was 15% for the Zooming Panorama and 14% for the Zooming Map. This shows how user performance with the zooming interfaces was less sensitive to the increase in task complexity. Yet, our results also show that the information provided by the Compass is sufficient if the system can highlight the target object: In real-world applications this could be the case if the system knows what the user is looking for (e.g. a result of a search query). The zooming interfaces support users better when browsing for information, in tasks where a more integrated view is needed: This would rather be the case when multiple options are available, and the selection criteria of a user are not clear or not easily described via a software interface.

We assumed to observe less increase in time and panning with the zooming interfaces than with the Compass, if information density increases (H3). We did not expect a difference between the two zooming interfaces. In the first study we found that high-density tasks took longer in general, independent of the interface used. In the second study we could not observe an effect of information density. Thus our data does not support our hypothesis.

In the first study we expected the Compass to require more panning than the other interfaces (H2). We indeed found that the Compass required significantly more movement than the Zooming Panorama, but there was no difference in the other interfaces.

During the two studies, we observed that the tilting movement used for triggering the Zooming Map can conflict with other actions that users normally perform. If some information is rather low on the vertical axis (e.g., a café in the basement) the users must tilt the phone down to see such information thus triggering an undesired Zooming Map transition. We also observed that some users tilt the phone down when looking at the Zooming Panorama visualization, probably because of outdoor lighting conditions (e.g. glare on the screen) since the tilted-down position helps blocking reflections with the arms. This highlights some limitations of our gesture-based approach: For tasks where the phone must be tilted down (e.g., visualizing underground pipes) such interaction would not be an appropriate solution.

8. CONCLUSION

While AR browsers can provide a straightforward mapping between real and virtual, the presented work highlights how two constraints – a fixed field of view and the first-person perspective – limit the amount of context information that can be provided. Instead of using an AR visualization for the entire interface, we have shown that integrating AR and other visualizations can ease search tasks in space.

While we foresee navigation guidance as a natural future step for this work, we are aware of some limitations of the study that do not allow us to generalize our findings to navigation tasks:

- We investigated search tasks from a static user position, implementing our prototype as a "point-and-click" system. Our interfaces did not allow users to move towards the target location, neither did the interfaces tell users how to get there.
- The locations of POIs were randomly generated and shuffled, to prevent learning effects. This did not provide users with a full matching between real and virtual: the information was bound to specific locations but not to real-world objects.
- Since the locations were continuously shuffled, we could not study how our interfaces supported users in the creation of spatial knowledge.

To work out these limitations we will follow our iterative approach, intermixing redesign stages and user involvement. Given the cross-subject nature of AR, our future work on AR browsers will necessarily touch different subject areas: data retrieval (e.g., retrieving POIs from online databases), tracking (e.g., improved sensor fusion), interaction and visualization (e.g., improved labeling algorithm, visual encoding of occlusion and distance of the POIs). With the next interface redesign we also plan to study how our interfaces can affect and support the development of spatial knowledge.

Finally, from our results we were not able to discern whether there are situations in which the Zooming Panorama outperforms the Zooming Map, and vice versa. Differences might depend on the type of task and information, as well as on subjective factors, and we plan to investigate on this in our future work.

9. ACKNOWLEDGMENTS

We would like to thank all users participating in the experiments. We also thank Raphaël Grasset for the feedback on early versions of the interfaces. This work was partially supported through the EU funded project MARCUS (FP7–PEOPLE–IRSES–2008-230831), through the Christian Doppler Laboratory for Handheld Augmented Reality and through the Austrian Science Fund FWF (W1209).

10. REFERENCES

[1] Avery, B., Sandor, C., and Thomas, B. Improving Spatial Perception for Augmented Reality X-Ray Vision, *In IEEE Virtual Reality*, pp. 79-82. Louisiana, USA. 2009.

[2] Billinghurst, M., Kato, H., and Poupyrev, I. 2001. The MagicBook—Moving Seamlessly between Reality and Virtuality. *IEEE Computer Graphics and Applications* 21, 3 (May. 2001), 6-8.

[3] Biocca, F., Tang, A., Owen, C., and Xiao, F. 2006. Attention funnel: omnidirectional 3D cursor for mobile augmented reality platforms. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Montréal, Québec, Canada, April 22 - 27, 2006). R. Grinter, T. Rodden, P. Aoki, E. Cutrell, R. Jeffries, and G. Olson, Eds. CHI '06. ACM, New York, NY, 1115-1122.

[4] Burigat, S., Chittaro, L., and Gabrielli, S. Navigation techniques for small-screen devices: An evaluation on maps and web pages. Int. J. Hum.-Comput. Stud. 66, 2, pp. 78-97, Feb. 2008.

[5] Büring, T., Gerken, J., and Reiterer, H. User Interaction with Scatterplots on Small Screens - A Comparative Evaluation of Geometric-Semantic Zoom and Fisheye Distortion. IEEE Transactions on Visualization and Computer Graphics 12, 5, pp. 829-836, Sep. 2006.

[6] Chittaro, L. 2006. Visualizing Information on Mobile Devices. *Computer* 39, 3 (Mar. 2006), 40-45.

[7] Cockburn, A., Karlson, A., and Bederson, B. B. A review of overview+detail, zooming, and focus+context interfaces. *ACM Comput. Surv.* 41, 1 (Dec. 2008), 1-31. 2008.

[8] Egenhofer, M. J. Spatial Information Appliances: A Next Generation of Geographic Information Systems. *1st Brazilian Workshop on GeoInformatics*, 1999.

[9] Feiner, S., MacIntyre, B., Hollerer, T., and Webster, T. A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. *In Proceedings of International Symposium on Wearable Computers (ISWC '97)*, pp. 74-81. 1997.

[10] Fröhlich, P., Simon, R., Baillie, L., and Anegg, H. 2006. Comparing conceptual designs for mobile access to geo-spatial information. *In Proceedings of the 8th Conference on Human-Computer interaction with Mobile Devices and Services (Helsinki, Finland, September 12 - 15, 2006). MobileHCI '06*, vol. 159. ACM, New York, NY, 109-112.

[11] Fröhlich, P., Obernberger, G., Simon, R., and Reichl, P. 2008. Exploring the design space of Smart Horizons. In *Proceedings of the 10th international Conference on Human Computer interaction with Mobile Devices and Services* (Amsterdam, The Netherlands, September 02 - 05, 2008). MobileHCI '08. ACM, New York, NY, 363-366.

[12] Grasset, R., Looser, J., and Billinghurst, M. Transitional interface: concept, issues and framework. In Proceedings of the Fifth IEEE and ACM international Symposium on Mixed and Augmented Reality (ISMAR'06). IEEE Computer Society, Washington, DC, 231-232. 2006.

[13] Kalman, R.E. A new approach to linear filtering and prediction problems, Trans. ASME, J. Basic Eng., vol. 82D, no. 1, pp. 34-45. 1960.

[14] Kooper, R., Macintyre, B. Browsing the real-world wide web: Maintaining awareness of virtual information in an AR information space. International Journal of Human-Computer Interaction, 16(3), 425-446. 2003.

[15] Lehikoinen, J. and Suomela, R. Accessing Context in Wearable Computers. *Personal Ubiquitous Comput.* 6, 1, pp. 64-74, Jan. 2002.

[16] Rekimoto, J. and Nagao, K. The world through the computer: computer augmented interaction with real world environments. *In Proceedings of the 8th Annual ACM Symposium on User interface and Software Technology* (UIST '95). ACM, New York, NY, 29-36. 1995.

[17] Robertson, G. G., Mackinlay, J. D., and Card, S. K. Cone Trees: animated 3D visualizations of hierarchical information. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Reaching Through Technology* (April 27 - May 02, 1991). ACM, New York, NY, 189-194. 1991.

[18] Schwerdtfeger, B. and Klinker, G. 2008. Supporting order picking with Augmented Reality. In *Proceedings of the 7th IEEE/ACM international Symposium on Mixed and Augmented Reality* (September 15 - 18, 2008). Symposium on Mixed and Augmented Reality. IEEE Computer Society, Washington, DC, 91-94.

[19] Spohrer, J. C. Information in places. *IBM Systems Journal* 38, 4, 602-628. 1999.

[20] Wagner, D., Reitmayr, G., Mulloni, A., Drummond, T., Schmalstieg, D., Pose Tracking from Natural Features on Mobile Phones, *In Proceedings of IEEE ISMAR'08*, Cambridge, UK. 2008.

[21] Zanella, A., Carpendale, M. S., and Rounding, M. On the effects of viewing cues in comprehending distortions. *In Proceedings of the Second Nordic Conference on Human-Computer interaction* (Aarhus, Denmark, October 19 - 23, 2002). NordiCHI '02, vol. 31. ACM, New York, NY, 119-128. 2002.