Map Navigation with Mobile Devices: Virtual versus Physical Movement with and without Visual Context

Michael Rohs
Deutsche Telekom Laboratories
TU Berlin, Germany
michael.rohs@telekom.de

Johannes Schöning
Institute for Geoinformatics
University of Münster, Germany
j.schoening@uni-muenster.de

Martin Raubal
Department of Geography
University of California at Santa Barbara, USA
raubal@geog.ucsb.edu

Georg Essl
Deutsche Telekom Laboratories
TU Berlin, Germany
georg.essl@telekom.de

Antonio Krüger
Institute for Geoinformatics
University of Münster, Germany
antonio.krueger@uni-muenster.de

ABSTRACT

A user study was conducted to compare the performance of three methods for map navigation with mobile devices. These methods are joystick navigation, the dynamic peephole method without visual context, and the magic lens paradigm using external visual context. The joystick method is the familiar scrolling and panning of a virtual map keeping the device itself static. In the dynamic peephole method the device is moved and the map is fixed with respect to an external frame of reference, but no visual information is present outside the device's display. The magic lens method augments an external content with graphical overlays, hence providing visual context outside the device display. Here too motion of the device serves to steer navigation. We compare these methods in a study measuring user performance, motion patterns, and subjective preference via questionnaires. The study demonstrates the advantage of dynamic peephole and magic lens interaction over joystick interaction in terms of search time and degree of exploration of the search space.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*input devices and strategies, interaction styles*

General Terms

Design, Experimentation, Human Factors

Keywords

maps, navigation, mobile devices, camera phones, interaction techniques, camera-based interaction, handheld displays, spatially aware displays, augmented reality

ICMI'07, November 12-15, 2007, Nagoya, Aichi, Japan.

1. INTRODUCTION

Map navigation with handheld devices helps mobile users to understand and explore their current place. But maps shown on handheld devices suffer from small display size and low resolution. It is often difficult to identify locations and landmarks on these maps. Traditional scrolling and panning interfaces with joystick or touch screen input offer only limited support in exploring large-scale maps on small displays. We investigate alternative map navigation techniques that are based on the *dynamic peephole* [9] and the *magic lens* metaphors [2].

In many mid- to large-sized cities, public maps are ubiquitous. They provide information and orientation not only to tourists, but also to locals who want to explore unfamiliar places. These maps are usually designed to address the most common questions of average users and therefore contain only general long-term information, such as street names and places of interest. More specific information, such as the locations of ATMs, pubs, shops, and restaurants is typically omitted for reasons of map complexity. The visualization of too many map elements eventually results in visual clutter and makes map interpretation difficult – if not impossible – for the average user [14].

Our implementation of a dynamic peephole interface consists of a camera-equipped mobile device that tracks its position above a marker grid. The physical position and distance of the device relative to the marker grid are used to pan and zoom to the corresponding location of the map shown on the display. Our magic lens approach to mobile map interaction takes advantage of the fact that geographic maps are typically highly structured and applies image analysis to track device position and orientation. By using their mobile phones as see-through tools, users can explore geospatial information on the map (see Figure 1). The main difference between the dynamic peephole and the magic lens interface is that the latter provides visual context outside the device display, while the former does not.

A user study investigates differences in task performance for mobile map navigation by contrasting the traditional static peephole interface, the dynamic peephole interface, and the magic lens interface. The general scenario for all three map navigation methods was to find an object with

Figure 1: Application example (constructed).

a specific attribute on the map. The results of this study demonstrate that users' search times are significantly reduced and the degree of exploration of the search space is significantly higher for dynamic compared to static peephole navigation. For the task and the map size studied there was no significant difference between the dynamic peephole interface and the magic lens, i.e., conditions with and without visual context.

2. BACKGROUND AND RELATED WORK

Various kinds of maps and other geospatial content can be found on the Internet today. With the growth of the mobile Internet, these geospatial data can be easily retrieved from mobile devices and, utilizing location based services (LBS), filtered based on position. Therefore, it is now possible to use geospatial data on the fly without a desktop computer. Gartner et al. [4] and Urquhart et al. [20] provide an overview of the major LBS prototypes that allow users to display maps and interact with them on a mobile device through ordinary interaction techniques, i.e., joystick or key input. *Google Maps Mobile*[1] is a prominent commercial example of a mobile geospatial service.

It is a natural problem to attempt merging the rich geospatial content available to augment existing physical maps. Mobile devices can serve as hosts for the additional information when interacting with a given map. Some approaches add hyperlink information to locations of a physical map. The mobile device can then display specific points of interest while the physical map provides a static overview. For example, Reilly et al. [11] use maps equipped with an array of RFID (Radio Frequency Identification) tags to realize the physical hyperlink. Disadvantages of this approach are the low spatial resolution (because of the size of the RFID tags) and the high map production costs. The method was enhanced by computer vision techniques [12] to achieve higher spatial resolution.

The term *magic lens* was coined to describe the use of mobile devices to augment a physical background in analogy to a reading glass. In a mobile augmented reality application the magic lens is a camera-equipped handheld device that mediates and enhances the user's view of the real world [13].

Applying the principle of magic lenses to paper maps is natural when using a mobile device with an integrated camera, such as a smartphone or PDA. When positioned over the map at a certain distance, map features recorded by the camera can be perceived on the display in real time. Following the classical video see-through augmented reality approach, digital geospatial data can then be overlaid over the images, delivering information to the user originally not available on the paper map. Geospatial data can be retrieved via a Web Map or Web Feature Service. Both are standardized protocols that allow requesting geospatial data across the web. This allows users to personalize each paper map with content of their interest without having the map cluttered with too much information (see Figure 1).

The magic lens metaphor is related to various forms of navigating virtual information in the plane. Mehra et al. [9] compare the *dynamic* and *static peephole* metaphors, which will be explained in detail in the following sections. They extend the work of Fitzmaurice [3] and Yee [21], and provide empirical evidence that a dynamic peephole interface is superior for tasks in which spatial relationships are important and the display size is limited. Mehra et al. use a mouse-based PC interface and focus on line-length discrimination whereas we focus on map exploration, target localization, and remembering a specific target attribute, hence requiring spatial memorization. The physical context of the interaction is different, since motor control issues of thumb movement vs. arm movement in 3D space play a role. Baudisch et al. [1] investigate the use of a high resolution focus display in combination with a lower resolution context display. Hachet et al. [6] realize a two-handed magic lens interface by tracking a piece of cardboard that the user moves behind a camera-equipped device. Sanneblad and Holmquist [17] use ultrasonic tracking to align a small display with a large overview in the context of a map application.

In our previous work [18] we have used a marker-based approach with an UMPC (Ultra Mobile PC) and a Symbian smartphone. The marker-based tracking approach has the severe drawback of requiring markers of considerable size scattered all over the map and thus disturbing its appearance, aesthetics, and overall usability, but the advantage that geospatial data can be precisely augmented on the map. Wikeye [7] is an approach to improve the understanding of places that combines digital Wikipedia content with a paper-based map. When the user views a small portion of a map through her mobile device, Wikipedia-derived content relating to these spatial objects is offered to the user.

3. IMPLEMENTATION OF THE NAVIGATION METHODS

Overall we have implemented three map navigation clients. The first and currently most common one is a simple reproduction of classical joystick-based map navigation applications. The two other methods are based on free physical movement of the mobile device over a grid of markers or a geographic map with a printed array of small dots. The marker grid does not provide visual context, whereas the map interface shows a static map in the background.

3.1 Static Peephole Navigation

Scrolling through a map via *joystick* input is the standard interaction technique available for mobile devices. Another common option is touch screen input. Joystick navigation is a static peephole interface [9]: the window is static while the

[1] http://www.google.de/gmm

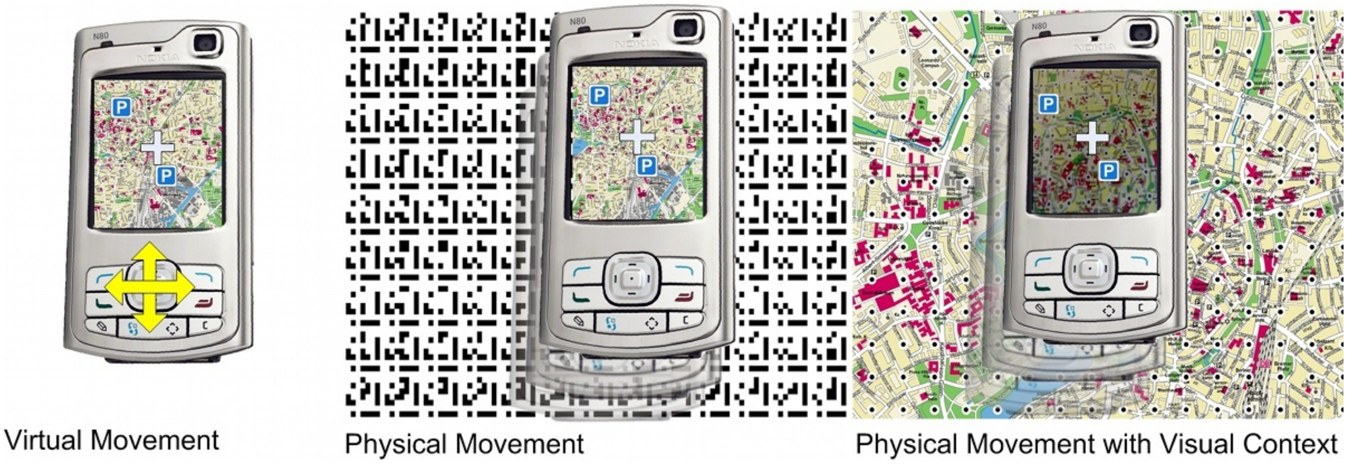

Figure 2: The investigated map navigation methods: in the static peephole interface the virtual map moves while the physical device is static (left); in the dynamic peephole interface (middle) and the magic lens interface (right) the physical device moves and the virtual map is static with respect to the background.

content moves behind it. The user must spatially and temporally integrate the movement of the virtual map. We used a device with a standard non-isometric 4-direction joystick. In addition, the center can be pressed to make a selection (see Figure 2, left).

In our implementation, when the joystick button is held in a direction, a timer generates update events at a rate of 13 Hz. At each timer event the display content scrolls by 40 pixels in the respective direction. This corresponds to 4 mm per update or 52 mm/sec on the testing device. The scrolling area was limited to the extent of the test map. For a map size of 1810×1280 pixels, as was used in the test described below, it thus takes 3.5 sec to scroll from left to right and 2.5 sec to scroll from top to bottom. The movement velocity was chosen as fast as possible, so that the features of the map could still be recognized during movement. The performance in the user study is thus limited by visual perception rather than scrolling speed. This is a general limitation of rate-controlled scrolling interfaces. If movement is too fast, users cannot observe the display or cannot effectively stop at the intended position.

3.2 Dynamic Peephole Navigation without Visual Context

In this navigation technique, the camera phone is tracked over a *grid* of visual markers. Aside from the handheld display no visual context of the map is available. The grid provides a fixed frame of reference for the virtual map on the handheld display. This setup is a dynamic peephole interface [9]: The map is fixed in space while the physical display is moved. The user has to temporally integrate the layout of the map. The grid defines a global coordinate system in which the camera phone can compute its (x,y,z) position precisely. The tracking method used here is an extension of the one described in [15]. The markers have been extended to a capacity of 16 bits: 2×7 bits for index positions and 2 parity bits. The maximum grid size is thus 128×128 markers or 1024×1024 code coordinate units, which enables large tracking areas (see Figure 2, middle).

In the original implementation the maximum tracking range (the distance of the camera lens to the grid surface) was limited to about 10 cm. This proved insufficient for effective interactions along the z-dimension. In particular, the range was too small for mapping to the zoom scale, because slight distance changes resulted in very rapid zoom scale changes. To extend the vertical tracking range, we use the digital zoom feature that is available in many camera phone APIs. Digital zoom increases the apparent focal length at which an image was taken by cropping an area at the image center with the same aspect ratio as the original image. The cropped area is rescaled to the original dimensions by interpolation. No optical resolution is gained in this process, but digital zoom is done by the camera and does not have to be performed by the main processor of the device.

In a pre-experiment, we kept the distance to an object in the camera view constant, continuously changed the digital zoom level, and measured the size at which the object appeared in the camera view ($size_{zoomed}$). We found a good fit of the measured data to $size_{zoomed} = size_{unzoomed} \exp(k\,level)$. For the Nokia N80 (20× digital zoom), which we used in the study described below, the constant k was determined as $k = 0.0345$ ($R^2 = 0.997$). With this constant and the above formula the unzoomed distance can be computed given the current zoom level.

During grid tracking, digital zoom is periodically adjusted, such that the markers appear at a size best suited for detection. The algorithm is complicated by the fact that changes to the zoom level via the camera API are not reflected in immediate changes in the next camera frame. Instead, the new digital zoom setting becomes valid with a delay of 2 to 5 frames after the adjustment is made. Therefore, the algorithm computes the unzoomed distance at the old and the new digital zoom levels and chooses the setting that yields the smoothest distance curve.

With this method, the vertical recognition range for a grid with a cell size of 1.5 mm is increased from 10 cm to about 50 cm. The grid interface scales the map shown on the display in real-time depending on the distance of the camera lens to the grid by resampling the original map image. Moving away from the grid is translated to zooming out, moving closer to the grid to zooming in. This is consistent with the behavior of the visual context interface described next.

3.3 Magic Lens Navigation with Visual Context

In this navigation method the camera phone is tracked over a static map, which at the same time provides visual context to the user. The user has complete overview of the static information available on the map. However, the dynamic information is only shown on the device display and the user has to switch attention between the device display and the map to benefit from the static overview. The approach currently requires an array of small black dots printed on the map (see Figure 2, right). An alternative would be horizontal and vertical lines commonly found on city maps.

The tracking algorithm we use is an improvement of [16]. It computes the position of the camera view on the map in real-time, so that appropriate graphical overlays can be generated with pixel-level accuracy. Since objects on the paper map may appear perspectively distorted in the camera view, this requires (for each camera frame) the computation of a projective mapping (planar homography) from the map coordinate system to the image coordinate system. The algorithm performs the following steps:

- *Find map dot candidates:* In the thresholded image, connected regions of a certain size and axis ratio are classified as potential map dots.
- *Find edges:* An undirected graph with the map dot candidates as vertices is computed and stored in a hashtable for efficient lookup.
- *Find patches:* The map dots subdivide the map into squared areas with map dots as corners (correlation patches). This step identifies the four corners of each correlation patch by iterating over the hashtable and looking for suitable edges.
- *Sample each patch:* For each patch a projective mapping to a 12×12 pixel area is computed and equally-spaced gray-value pixel samples are taken.
- *Compute correlations:* The correlation coefficients between each patch found in the image and each stored 12×12 pixel map patch are computed. This step is the most computationally intensive and was thus optimized as much as possible.
- *Compute maximum correlation indices:* For each patch in the camera image the best correlating map patch is determined. The result is a list of pairs of image patches and map patches.
- *Find reliable patch:* A voting scheme is used to identify patches whose position on the map has been correctly recognized. If two patches agree in their prediction of the cursor position, they are classified as reliable.
- *Set map dot coordinates:* Since each individual reliable patch might be too small to provide a stable projective mapping, graphical overlays are based on a larger area. A reliable patch serves as a seed to infer the exact coordinates of the other map dots in the camera image.
- *Compute maximum warper:* Finally, the perspective mapping for the graphical overlays is based on the map dots closest to the image corners.

For performance reasons the interface shows the camera view of the paper map in gray-scale and overlays the dynamic information on top of the camera image. In addition, only a window of 5×5 patches around the previous position is considered for correlation to increase performance.

4. USER STUDY

We conducted a user study to compare the three different map navigation techniques presented above. Our initial hypothesis was that dynamic peephole and magic lens navigation would outperform the static peephole navigation method (*joystick*), and that the magic lens (*map*) interface with visual context would outperform the dynamic peephole interface without visual context (*grid*) (see Figure 2). The task was chosen such that it combines static long-term information on the map with dynamic online information. In all cases, the dynamic information was only available on the device display.

We chose to augment parking lot symbols on a city map of Münster, Germany, with hourly rates for parking. The task was to find the cheapest parking lot on the map. Even though this task might be easily solved automatically, rather than letting the user do an extensive search, many tasks exist where the user does not know in advance which option she will finally choose. Also, it may not easily be possible to effectively tell the system her preferences. In such cases explorative map navigation is beneficial. The advantage of the used task is that the cheapest rate can easily be reassigned to different parking lots and a large number of test cases can automatically be generated this way. Finding the cheapest parking lot requires subjects to:

- examine all candidate objects in turn,
- keep the location and amount of the best one in mind,
- determine when the whole space was explored, and
- navigate back to the memorized target;

The study thus combines small display navigation and spatial memory components.

4.1 Participants and Apparatus

The study was conducted with 18 participants, 10 female, 8 male, ages 21-33. The subjects were undergraduates, doctoral students, or post-doctoral researchers with varying degrees of technical background. None of the subjects was familiar with the city or the map (see Figure 2, right).

The test was performed on a Nokia N80 Symbian phone. 352×288 pixels (35×29 mm) of the total display area (35 × 41 mm) were used to show the map. Two identical devices with this setup were used. During each trial, the (x, y, z) coordinates ((x, y) for *joystick*) of the motion trajectory were sampled at an average rate of 7.2 Hz: Holding the non-isometric 5-way joystick button in one of the directions generated 13 updates per second. Since subjects repeatedly released the button to view the display contents, the effective update rate was 8.0 Hz. The *grid* method provided 7.0 Hz and the map method achieved 6.5 Hz. The time to target selection for each trial and the success or failure of a trial were recorded. The dimensions of the map in the virtual workspace were 1810×1280 pixels.

For the physical movement condition without visual context a black-and-white marker grid printed on an A3 sheet was attached to the wall. For the physical movement condition with visual context, the city map printed in color on an A3 sheet was attached to the wall. For each user the height of the sheets was adjusted such that the upper border was at eye height.

The *joystick* interface shows a portion of the map at close-up view. The *grid* interface scales the map depending on the distance. The operable distance range of the *map* interface is 6-21 cm. The recognition rate is quite uniform for *grid* because of the dynamic digital zoom feature and drops significantly for *map* as one approaches the limits of the recognition range. To give feedback about the distance limits, the text "too close" and "too far," respectively, was displayed when users left the recognition range. For the *grid* method, a maximum distance for showing the rates was set in order to provide a distance performance comparable to the *map* interface. This distance was chosen such that the size of the zoomed area visible on the screen was identical to that visible with the *map* technique at the boundary of the recognition range.

4.2 Tasks

The general scenario for all three map navigation methods was the same. Users had to find the cheapest among 13 parking lots on the map. Each parking lot was marked with a blue P symbol. There was always a unique cheapest target present on the map. The cheapest rate varied randomly between €0.50 and €1.20. Increments were €0.10 or €0.20. Duplicates were possible (except for the cheapest rate). The rates were randomly assigned to P symbols and displayed in red with a black shadow below each P symbol.

A single trial consisted of navigating the map using the given method and finally selecting the target. At any time the P symbol closest to the cursor on the screen's center was highlighted with a red frame. We did not require users to exactly locate the cursor on the target, since we are not focusing on a Fitts' law task, but on mobile map navigation. After each selection the subject was informed about success or failure of the trial and the next trial could be started. 15 trials were done for each condition.

4.3 Design

The study was set up as a within-participants design with the map navigation method as the single factor.

- *Joystick*: static peephole navigation on the display only.
- *Grid*: dynamic peephole navigation with a spatially tracked display and visualization on the display only.
- *Map*: magic lens navigation with a spatially tracked display. Visualization on the display (camera image plus overlays) and static visual context with the paper map on the wall.

The order of navigation methods was counterbalanced and presented in blocks. For example, all *joystick* interactions happened in one block without allowing the user to switch to another method. The assignment of rates to each P symbol within the trials was randomized as described above. With 15 trials per method, 3 methods per subject, and 18 subjects the test application on the devices recorded 806 trials and a total of 177,489 cursor events. Four trials were lost due to participants accidentally exiting the test application.

4.4 Procedure

Initially, participants were given a written task description. Next, each method was briefly demonstrated and the limits of the recognition range shown. The height of the map and the grid sheet were adjusted to the subject's body height. This initial phase took about 10 to 15 minutes. The participants performed the methods in the order given by the test application. There were no practice trials before the actual test. After each trial, there was a pause screen that informed the user about the number of completed trials in this block and the current navigation method. Subjects were requested not to talk during trials but only when the pause screen appeared. When the participants were ready, they clicked the right selection button on the camera phone to start the next trial. Target selection was done with the center joystick button. After the actual test users were asked to rate the map navigation techniques by filling out a modified version of the "user interface evaluation questionnaire" of ISO 9241-9 [8] with only a single Fatigue category. The ISO questionnaire is a seven-point rating evaluation. Higher scores denote a better rating. The total time each participant took was about 50 minutes.

4.5 Results

All participants were able to complete the task for all methods. The main performance measures taken are trial time and error rate. Trial time is the time from the start of a trial until the selection is made. The error rate is the ratio of selections made on another than the cheapest parking lot. The overall average trial time is 31.7 sec (95% confidence interval: 30.8-32.6 sec) and the overall average error rate is 15% (95% confidence interval: 9-21%). In these numbers, outliers more than 2 standard deviations from the mean and the first trial of each navigation method are excluded. Figure 3 shows (from left to right) the average trial times, error rates, target omission ratios, and viewing times per target by navigation method. With 29.4 sec *grid* is 23% faster and with 28.3 sec *map* is 26% faster than *joystick* (38.4 sec). An analysis of variance (ANOVA) on trial times shows a significant effect of navigation method ($F(2,720) = 56.15$, $p < 0.001$). Pairwise tests reveal that the differences between *joystick* and *grid* ($F(1,479) = 65.21$, $p < 0.001$) and between *joystick* and *map* ($F(1,477) = 92.25$, $p < 0.001$) are significant, but surprisingly the difference between *map* and *grid* is not ($F(1,484) = 1.41$, $p = 0.236$). The differences in error rate are within the limits of the 95% confidence interval and thus not significant at the 5% level. We also examined the coverage of target candidates, i.e., the number of targets that were visited and present on the display, in order to estimate the degree of exploration of the search space. Candidate coverage was significantly lower for *joystick* (82.5%) than for *grid* (89.2%) and *map* (88.7%) ($F(2,747) = 11.98$, $p < 0.001$). Coverage differences between *map* and *grid* were not significant ($F(1,499) = 0.52$, $p = 0.4712$).

A noticeable learning effect was only present for the first trial of each method (see Figure 4). In the ISO questionnaire *joystick* was rated best for force, smoothness, accuracy, overall operation, and fatigue. *Grid* was rated best for effort, speed, and comfort. However, the answers varied strongly, which is reflected in the large confidence intervals (see Figure 5). None of the differences is significant at the 5% level.

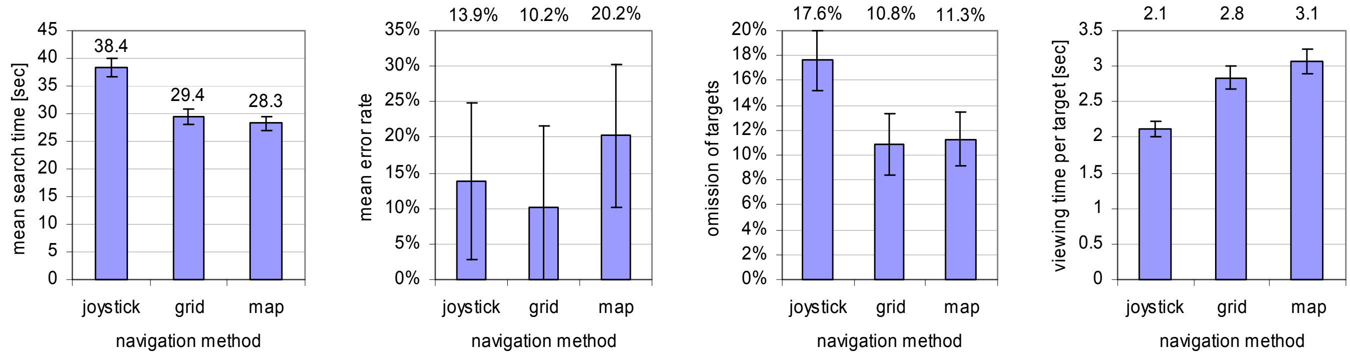

Figure 3: Mean search times for each map navigation method, mean error rates, ratios of omitted targets, and viewing times per target. Error bars in all figures show 95% confidence intervals.

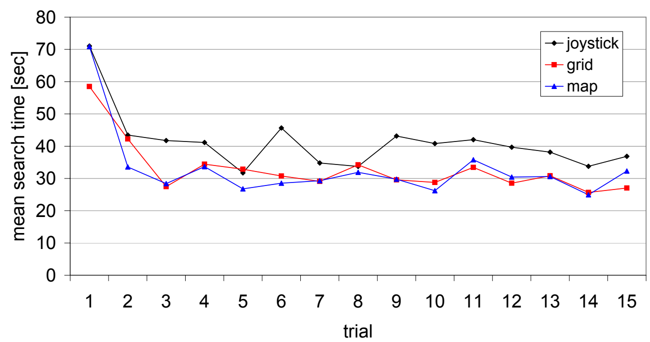

Figure 4: Learning effects by trial and navigation method.

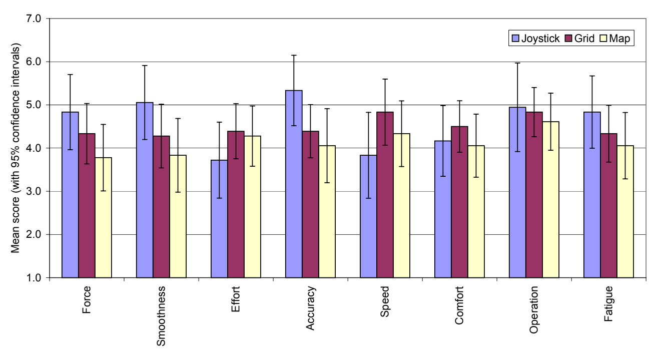

Figure 5: Results of the user interface evaluation questionnaire.

4.6 Motion and Search Strategies

Figures 6, 7, and 8 demonstrate typical traces of single trials for the different navigation methods. All traces show successful selections and are from the first block, i.e., user behavior is not influenced by preceding navigation techniques. The green dot indicates the start of the trial, the red dot denotes the selection position. Each P symbol is labeled with the duration of visibility on the device display. Symbols that were never visited have a value of 0.0 sec and are highlighted with a red frame. In the *joystick* condition depicted in Figure 6 two P symbols were never examined (visit time 0.0 sec). Each point on the trace represents a position update. Closely spaced points thus denote slow movement. The *joystick* trace starts with a movement down to the map borders and then continues with a more or less systematic exploration of the space. In the *grid* condition the user performs a counter clockwise circular movement of nearly constant velocity (see Figure 7). The velocity in the *map* condition varies much more (see Figure 8). The user moves quickly from symbol to symbol and does not lose time in-between. The movement is strongly influenced by the positions of the P symbols, therefore visual context appears to help performing the task. The participant seems to rely on the visual context to quickly acquire the next target.

4.7 Discussion

The dynamic peephole and the magic lens navigation methods clearly outperform the static navigation method. However, other than expected *map* did not turn out to be substantially faster than *grid*, even though *map* provides static visual context in a second layer of information. We found two potential reasons that might account for this result, one rooted in system performance, the other in human visual perception.

On the technical side, the implementation of *map* has a slightly lower update rate and lower recognition reliability. The recognition reliability of *grid* is quite insensitive to the distance between the marker surface and the lens, because of the dynamic digital zoom feature. In contrast, *map* has a more restricted distance range and recognition reliability decreases as one approaches the limits of the recognition range. The dynamic zoom feature could in principle also be included in the *map* interface, but it might be distracting if the zoom level automatically follows the movement and always keeps the objects in the camera view at a constant size. An alternative would be to not show the camera image, but a continuously scaled virtual version of the map, as in the *grid* technique. This would have the additional advantage that the display quality does not depend on the quality and resolution of the camera image.

In terms of human visual perception, switching between the two layers of visual presentation as is necessary with the *map* interface might incur higher costs than expected. With each switch of layers the user's eyes have to refocus on the new depth and locate the intended object on the new layer. Since the P symbols on the map were relatively large and had a different color from the surrounding area they could easily be spotted at a glance and even if slightly out of focus. For smaller elements this effect might be more severe.

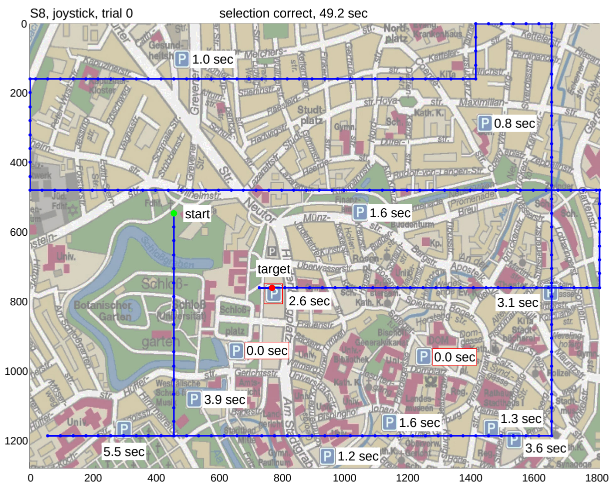

Figure 6: Typical trace for a *joystick* trial.

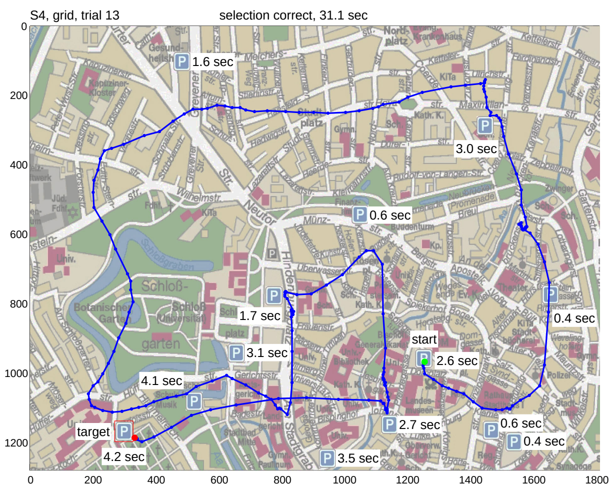

Figure 7: Typical trace for a *grid* trial.

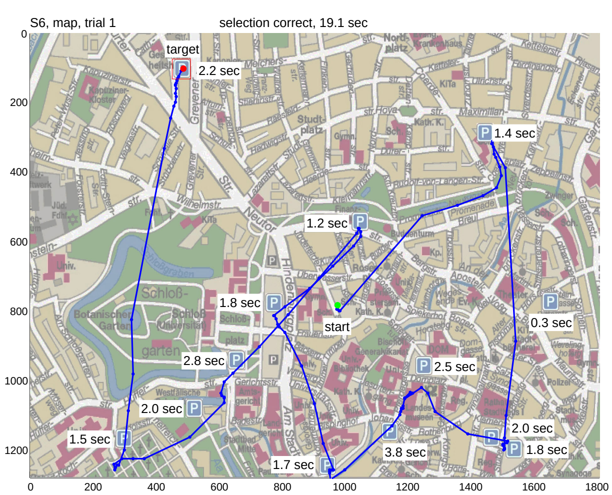

Figure 8: Typical trace for a *map* trial.

Another issue is that the frame of the device occludes part of the background layer. This is not a problem for moving towards a target that is far from the previous location, but for moving within clusters of objects being densely spaced. In such cases the user has to look around the handheld device in order to see the background. Larger map sizes, which would spread the map's elements over a larger area, might be beneficial to reduce occlusions. Even though we told the participants in the beginning of the test that they should use the paper map in combination with the mobile device, a few concentrated on the mobile display without ever looking at the background. This behavior might be due to the unfamiliarity of the subjects with hybrid interaction methods.

These results are of course strongly influenced by the chosen task, which required users to look onto the mobile device display to see the fees for the parking lots. The background only provided information about the locations of the parking lots. Other tasks might allow gaining more information from the static map in the visual context, which should reduce the time needed for interaction with the device. In such a situation the *map* interface is expected to show clear advantages in comparison to the *grid* technique.

5. CONCLUSIONS & FURTHER WORK

This work compares three methods for navigating in geographic maps using mobile devices: a traditional joystick-controlled interface and two interfaces based on visual tracking. The first is a dynamic peephole interface without visual context. It is implemented by tracking a camera-equipped mobile phone over a marker grid and enables access to different parts of a map through physical movement. The second follows a magic lens approach that provides seamless real-time tracking over a printed map and augments the focused part with graphical overlays. The main contribution is the investigation of mobile map navigation comparing virtual and physical movement and the impact of visual context. The comparison was done in an unobtrusive way with off-the-shelf camera phones.

The results demonstrate that map exploration performance is better when moving a device over a static map, than when moving a virtual map behind a static device. In addition, more of the map space is explored in the device-tracking condition. The results also indicate that for the task and map size chosen providing visual context does not automatically lead to performance improvements. Switching visual attention seems to incur some cost. We will explore this interesting result in more detail in a separate study in the future. The chosen task involved simple spatial exploration and required navigation to all candidate objects on the map and inspecting them through the device display. The magic lens approach is expected to show a much better performance if more information can be acquired from the static map without requiring mediation by the device.

In the magic lens condition the small dots seemed to be unobtrusive; many participants did not even notice them. They can be replaced by grid lines commonly found on many city maps. The *grid* method has been used in earlier studies and has matured and gained robustness in the process. The *map* method was newly developed.

In the future, we want to increase the numbers of modalities. Haptic feedback can be added to give feedback if users cross the mobile device over a P symbol. Such feedback could help users to select targets more accurately. Map sonification is a well studied technique. Effective map sonification can help vision-impaired users to explore topographic data for problem solving and decision making. Adding more modalities to our approach for exploring a map with a multimodal interface provides new user interface metaphors that hold potential for a wide range of users, such as people who have problems reading a map or visually impaired persons.

Besides adding more modalities another question to be pursued in the future is that of people's spatial knowledge acquisition [5] through mobile map interaction. When exploring an unfamiliar environment, a major spatial cognition task consists of acquiring different types of spatial knowledge, such as landmark, route, and survey knowledge [10, 19]. It needs to be investigated, which of the navigation strategies facilitates such knowledge acquisition and why. One possible method is to have subjects draw sketch maps of the presented environment and analyze them with regard to their elements. A possibly important factor for the study results regarding the *map* navigation technique is how often users switch between the paper map and the mobile display. A large number of such switches could lead to a better overall understanding of the environment but might also lead to higher distraction and therefore increased task complexity. The use of an eye-tracking apparatus would allow investigating such questions.

6. ACKNOWLEDGMENTS

We thank Gwen Raubal, Brent Hecht, and Michael Schellenbach for their help with improving the video, constructive feedback, and helping with the user study.

7. REFERENCES

[1] P. Baudisch, N. Good, V. Bellotti, and P. Schraedley. Keeping things in context: A comparative evaluation of focus plus context screens, overviews, and zooming. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 259–266, 2002.

[2] E. A. Bier, M. C. Stone, K. Pier, W. Buxton, and T. D. DeRose. Toolglass and magic lenses: The see-through interface. In *SIGGRAPH '93: Proc. of the 20th annual conf. on Computer Graphics and Interactive Techniques*, pages 73–80, 1993.

[3] G. W. Fitzmaurice. Situated information spaces and spatially aware palmtop computers. *Commun. ACM*, 36(7):39–49, 1993.

[4] G. Gartner, M. P. Peterson, and W. Cartwright. *Location Based Services and TeleCartography.* Springer Verlag, 2007.

[5] R. Golledge. Place recognition and wayfinding: Making sense of space. *Geoforum*, 23(2):199–214, 1992.

[6] M. Hachet, J. Pouderoux, P. Guitton, and J.-C. Gonzato. TangiMap: A tangible interface for visualization of large documents on handheld computers. In *GI '05: Proceedings of the Conference on Graphics Interface*, pages 9–15, 2005.

[7] B. Hecht, M. Rohs, J. Schöning, and A. Krüger. Wikeye – using magic lenses to explore georeferenced Wikipedia content. In *Proceedings of the 3rd International Workshop on Pervasive Mobile Interaction Devices (PERMID)*, 2007.

[8] International Organization for Standardization. Ergonomic requirements for office work with visual display terminals (VDTs) – Requirements for non-keyboard input devices. ISO/IEC 9241-9, 2000.

[9] S. Mehra, P. Werkhoven, and M. Worring. Navigating on handheld displays: Dynamic versus static peephole navigation. *ACM Trans. Comput.-Hum. Interact.*, 13(4):448–457, 2006.

[10] D. Montello. A new framework for understanding the acquisition of spatial knowledge in large-scale environments. In M. Egenhofer and R. Golledge, editors, *Spatial and Temporal Reasoning in Geographic Information Systems. Spatial Information Systems*, pages 143–154. Oxford Univ. Press, 1998.

[11] D. Reilly, M. Welsman-Dinelle, C. Bate, and K. Inkpen. Just point and click?: Using handhelds to interact with paper maps. In *Proceedings of MobileHCI*, pages 239–242. ACM Press, 2005.

[12] D. F. Reilly, M. E. Rodgers, R. Argue, M. Nunes, and K. Inkpen. Marked-up maps: Combining paper maps and electronic information resources. *Personal and Ubiquitous Computing*, pages 215–226, 2006.

[13] J. Rekimoto. The magnifying glass approach to augmented reality systems. In *International Conference on Artificial Reality and Tele-Existence / Conference on Virtual Reality Software and Technology (ICAT/VRST)*, pages 123–132, 1995.

[14] C. Rinner, M. Raubal, and B. Spigel. User interface design for location-based decision services. *13th Intl. Conf. on GeoInformatics, (CD-ROM)*, 2005.

[15] M. Rohs and G. Essl. Which One is Better? – Information Navigation Techniques for Spatially Aware Handheld Displays. In *Proc. of the Eighth International Conference on Multimodal Interfaces*, pages 100–107, November 2006.

[16] M. Rohs, J. Schöning, A. Krüger, and B. Hecht. Towards real-time markerless tracking of magic lenses on paper maps. In *Adjunct Proceedings of the 5th Intl. Conference on Pervasive Computing (Pervasive), Late Breaking Results*, pages 69–72, 2007.

[17] J. Sanneblad and L. E. Holmquist. Ubiquitous graphics: Combining hand-held and wall-size displays to interact with large images. In *AVI '06: Proceedings of the Working Conference on Advanced Visual Interfaces*, pages 373–377, 2006.

[18] J. Schöning, A. Krüger, and H. J. Müller. Interaction of mobile camera devices with physical maps. In *Adjunct Proceedings of the 4th Intl. Conference on Pervasive Computing (Pervasive)*, pages 121–124, 2006.

[19] A. Siegel and S. White. The development of spatial representations of large-scale environments. In H. Reese, editor, *Advances in Child Development and Behavior*, volume 10, pages 9–55. Academic Press, New York, USA, 1975.

[20] K. Urquhart, W. Cartwright, S. Miller, K. Mitchell, C. Quirion, and P. Benda. Cartographic representations in location-based services. *Proc. 21st Intl. Cartographic Conference*, pages 2591–2602, 2003.

[21] K.-P. Yee. Peephole displays: Pen interaction on spatially aware handheld computers. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 1–8, 2003.