Pers Ubiquit Comput (2009) 13:633–646
DOI 10.1007/s00779-009-0247-2

ORIGINAL ARTICLE

Impact of item density on the utility of visual context in magic lens interactions

**Michael Rohs · Robert Schleicher ·
Johannes Schöning · Georg Essl ·
Anja Naumann · Antonio Krüger**

Received: 17 October 2008 / Accepted: 15 March 2009 / Published online: 4 July 2009

Abstract This article reports on two user studies investigating the effect of visual context in handheld augmented reality interfaces. A dynamic peephole interface (without visual context beyond the device display) was compared to a magic lens interface (with video see-through augmentation of external visual context). The task was to explore items on a map and look for a specific attribute. We tested different sizes of visual context as well as different numbers of items per area, i.e. different item densities. Hand motion patterns and eye movements were recorded. We found that visual context is most effective for sparsely distributed items and gets less helpful with increasing item density. User performance in the magic lens case is generally better than in the dynamic peephole case, but approaches the performance of the latter the more densely the items are spaced. In all conditions, subjective feedback indicates that participants generally prefer visual context over the lack thereof. The insights gained from this study are relevant for designers of mobile AR and dynamic peephole interfaces, involving spatially tracked personal displays or combined personal and public displays, by suggesting when to use visual context.

Keywords Magic lens · Dynamic peephole · Small display · Mobile device · Camera phone · Focus and context display · Visual search · Eye tracking · Saccades · Pupil dilation

1 Introduction

Mobile devices provide a convenient way to augment existing static information with dynamic and personalized content. For example, large-scale paper maps are already available in public spaces but they only provide long-term, structural information that is intended for broad use by a general audience. Mobile devices can add specific content dynamically and hence increase the value of static large-scale maps for navigation and exploration. At the same time, mobile devices have limited screen space and do not provide a good overview over large visual information areas. Combining the advantages of large-scale paper maps and of small dynamic personal displays has the potential to overcome both problems [1, 2].

In our approach camera phones use the integrated camera to precisely track their position over a background surface and overlay additional information over the video stream in real time [3]. This so-called magic lens [4] approach allows users to use general public displays in a personalized way by dynamically adding selected content. For example, a standard city center map could provide selected points of interest, like nearby coffee places or

---

M. Rohs (✉) · R. Schleicher · G. Essl · A. Naumann
Deutsche Telekom Laboratories, TU Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Germany
e-mail: michael.rohs@telekom.de

R. Schleicher
e-mail: robert.schleicher@telekom.de

G. Essl
e-mail: georg.essl@telekom.de

A. Naumann
e-mail: anja.naumann@telekom.de

J. Schöning · A. Krüger
Institute for Geoinformatics, University of Münster, Weselerstrasse 253, 48151 Münster, Germany
e-mail: j.schoening@uni-muenster.de

A. Krüger
e-mail: antonio.krueger@uni-muenster.de

museums, by holding a camera phone over it. The camera phone can serve as an entry point to online information and services, such as current events, opening times, current movies at a cinema, ticket hotlines, navigation services, as well as parking lots with their associated parking costs (Fig. 1).

However, using two displays in combination potentially creates new usability challenges, because users have to divide their attention between the handheld device display and the large public display. Locating items of interest involves looking at the large background for identifying static features, as well as using the handheld display for getting up-to-date online information. The exploration of these two presentation surfaces requires visual search in each one and switching visual attention between the two. This paper tries to identify the strategies users adopt to solve typical search tasks in this context and to chart the performance that can be expected in such interactions.

A previous study on exploring maps with mobile devices [5] compared the performance of traditional joystick navigation (static peephole [6]), position-tracked navigation without visual context (dynamic peephole [6]), and position-tracked navigation with visual context (magic lens [4]). In the dynamic peephole case, the map information is only presented on the device display. In the magic lens case, map information is available on both the personal device display and the large-scale paper map. The magic lens provides video see-through augmentation of the external map. In this previous study [5], the two position-tracked interfaces outperformed the static peephole navigation method (joystick), but the magic lens interface (which provides visual context) surprisingly was not significantly faster than the dynamic peephole (which does not provide visual context). As the second experiment presented in this article shows, the lack of an advantage in the magic lens case was caused by the particular choice of item number and background size.

Fig. 1 Camera view augmented with prices for parking lots. The *blue parking signs* are visible on the background map. The overlay graphics are generated by the phone

The goal of this work is to explore the role of the size of the visual context and the item density in interactions, which combine a personal handheld and a large-scale context display. The first study presented below is explorative and investigates what movement patterns and gaze shifts occur. The second study investigates in more detail the role of display size and item density. Our initial hypothesis was that the item density would affect to what extent users take advantage of the information that is provided in the background. Specifically, we hypothesized that the effectiveness of visual context decreases as item density increases. We expected that for lower item densities the magic lens condition (with visual context) would outperform the dynamic peephole condition (without visual context). The earlier experiment [5] suggested that there is a density limit above which users will only use the device display and not switch their visual attention to the background display. The results of this study can help to decide whether it is useful to offer visual context in the background or just use a dynamic peephole interface.

2 Related work

Camera-equipped mobile devices can be used as see-through tools [4] to augment background surfaces, such as paper maps, posters, or electronic displays. When the device is held above an object or surface, visual features in the scene are highlighted and additional information is overlaid in real-time on the device's display (see Fig. 1). The term magic lens [4] has been coined in the context of graphical user interfaces to describe this type of multi-layer interface in analogy to a reading or magnifying glass [7].

Whereas magic lens interfaces are based on the idea of real-time augmentation of the real world scene, peephole interfaces [6, 8] denote a class of interface where the viewport of a mobile device is used as a window into a virtual space and no visual context is available outside the device display. This requires a spatial tracking method in order to compensate for the movement of the peephole, such that the workspace appears at a constant position in space [9]. As an example of a dynamic peephole interface, Yee [8] prototyped a spatially aware calendar application. Hachet et al. [10] realized a two-handed interface by tracking a piece of cardboard that the user moves behind a camera-equipped device.

Magic lens interfaces with external context offer a particularly promising kind of interaction, since they allow for augmenting large-scale public displays with high-resolution information on the handheld device. This relates to the concept of focus and context displays. Baudisch et al.

[11] investigated the use of a high resolution focus display in combination with a lower resolution context display. Sanneblad and Holmquist [12] used ultrasonic tracking to align a small display with a large overview for a map application.

Paper maps are highly structured ways of visual information presentation that particularly benefit from a large area in order to effectively depict spatial relationships. Several approaches exist to link paper maps to electronic content using handheld devices. For example, Reilly et al. [1] deployed maps equipped with an array of RFID tags to the realize physical hyperlinks. The number of hyperlinks is limited to the number of RFID tags used and the map production costs are relatively high. Schöning et al. [2] used a marker-based approach with a camera-equipped PDA to augment paper maps. The idea of Wikeye [13] is to place georeferenced Wikipedia content on public city maps in order to help users learn more about their current place. When the user views a small portion of the map through his or her mobile device, Wikipedia-derived content relating to these spatial objects is offered to the user.

In contrast to the more general work outlined above, we investigate users' search strategies in magic lens interaction and in particular the effects of item density and background size on the effectiveness of visual context. This enables us to formulate heuristics for the usage of a magic lens interfaces and to characterize the performance gain that can be expected by providing visual context in a given situation, e.g., for searching.

Several models of visual search have been proposed in psychology that differ in the extent of parallelism they assume, i.e. to what extent processes are supposed to occur in parallel or serially [14, 15]. In parallel search, the target "pops out" from distractor items as it is uniquely different (e.g. in color), and an increase of distractor items does not increase the search time, whereby in serial search, item detection requires detailed processing (e.g. the letter "C" in a group of "O"s), probably as the target item consists of a combination of different features (e.g. a red "C" in a group of red "O"s and green "C"s) that have to be evaluated in conjunction (Feature Integration Theory [16]). Here, the number of distractor items influences search time, although it might be possible that subjects group items if they know all relevant item features in advance (Guided Search Model [17]). The slope of this increase in search time is supposed to reflect search difficulty.

The task used in the studies presented below was to find the cheapest parking lot. This task is obviously an exhaustive serial search, as all items have to be checked to determine the cheapest among them. The first step in this task, to locate the P symbols, however, may be done in parallel if the map appears on the large background whereas it can only be achieved by serial scanning with the magic peephole if no meaningful information is presented on the large display and the map just appears on the mobile phone display. The question is to what extent the information in the background will be used during the task. Focusing on a nearby display is more demanding than watching one at a longer distance due to higher load on the ocular vergence system [18]. On the other hand, switching between levels of different distances also requires vergence movements as well as changes in lens contraction. With regard to scanning behavior, the next fixated item is the more likely to be hit correctly with a saccade, the closer it is to the current fixation point [19]. Large saccades appear to be associated with more planning costs and thus there is a tendency to prefer short saccades, especially if arm movements are involved [20]. Additionally, people show difficulties to guide attention away from the area near their hands when manual interaction is involved [21].

3 Overview of experiments

In the following, we report two experiments. Both involve searching on maps with the help of a mobile phone. In addition to subjective responses, reaction times, and error rates, we also recorded eye movements. However, the two experiments differ with regard to stimulus material and eye movement evaluation. In the first study, the number and position of items remained constant throughout the trials. In the second study, the position of items changed with each trial and the number of items varied as an additional factor. The deployed eye tracking analysis in the first study was restricted to manual coding of the video recording; in the second study it additionally involved a detailed analysis of saccadic and fixation parameters.

The first experiment was primarily explorative and was aimed to reveal strategies that users adopt when solving a basic search task with a combined handheld and static display. The movement trace of the phone across the background was logged on the device and eye movements were recorded via a head-mounted eye tracker. The eye movement videos were manually evaluated to capture shifts between the device display and the background, but no detailed analysis (e.g., of the duration of fixations) was performed. The first experiment provided valuable insight about the interactions, but had a few limitations. First, the study was performed with a set of static paper maps. Therefore, participants might have remembered the positions of the items, which would potentially influence results. Second, we noticed that the fixed number of 20 items on the map was quite high. Frequently, the next item already appeared on the device display, which made it unnecessary for the user to scan the background for the next item.

In the second study the map in the background and the positions of the items were changed after each trial. Moreover, the number of items was introduced as an additional factor. The analysis of eye movement behavior was more detailed. In addition to manual analysis of gaze shifts, we determined the average amplitude of saccades and maximum pupil dilations as an indicator of mental workload [22].

4 Experiment 1

The first study investigated the effects of the presence of visual context on completion time, error rate, and subjective satisfaction in a basic search task involving a small handheld and a large static display. In the magic lens case, the items were visible on the background, but the attribute to look for was only available as a textual overlay generated by the phone. The aim was to simulate searching for dynamic information that cannot be expected to be completely available on a static background.

4.1 Design

The first study was set up as a 2 × 2 within-participants design with the following factors (see Fig. 2):

1. Visual context: available (city map) versus not available (abstract pattern).
2. Context size: small (A3) versus large (A1).

The conditions without visual context (pattern) implied dynamic peephole navigation with the spatially tracked display and visualization on the display only. In contrast, the conditions with visual context (map) implied magic lens navigation, where the mobile display reproduced the underlying paper map section plus overlays (as in Fig. 1).

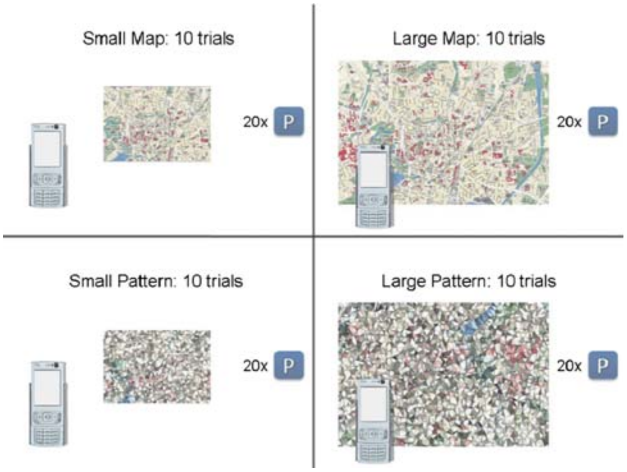

Fig. 2 Experimental conditions of experiment 1

The order of conditions was counterbalanced and presented in blocks. For example all small map interactions appeared 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. One block consisted of ten trials. Thus, participants had to complete 40 trials altogether.

4.2 Tasks

To cover a typical task for mobile map interaction we chose an object locator task, which is described as a fundamental task in the literature [23]. The general scenario for all conditions was that users had to find the cheapest among 20 parking lots on the map (indicated by blue parking signs). For the conditions with visual context the parking signs were visible in the background, but the price for parking was only visible on the phone (see Fig. 1).

A single trial consisted of scanning the map in the defined condition and finally selecting the target. At any time the item closest to the cursor (on the screen's center) was highlighted with a red frame (see Fig. 1) and selected when the user pressed a button. Users were not required to exactly position the cursor on the target. After each selection the participants were informed about success or failure of the trial and the next trial could be started. After finishing 10 trials per condition a screen informed the participants about the next condition.

4.3 Apparatus

For the condition without visual context an abstract colored pattern, generated with an image mosaic algorithm, was printed on paper sheets both in A3 2 × 2 × 4 and A1 size 2 × 2 × 4 and attached to a whiteboard in landscape orientation. For the condition with visual context a colored city map was printed in both sizes and attached to the whiteboard as well.

The handheld device was a Nokia N95 camera phone. The same real-time tracking method was used in both experiments and in all conditions [3]. It provides graphical overlays with pixel-level accuracy, has an average frame rate of 8–10 Hz, and a delay below 170 ms. In addition to the algorithm described in [3], fast movements are detected with an optical flow method. This provided sufficient responsiveness for our purpose.

The magic lens function was enabled by a client application that captured all user interactions and movements with timestamps. With this application it is possible to see the video stream of the phone camera on the phone display augmented with additional information. The operable distance range of the phone was 6–21 cm. To give feedback about the distance limits the text “too close” and “too far,” respectively, was displayed when users were about to leave the recognition range. On the display the city map with parking lots was shown. Each parking lot was marked with a blue “P” symbol. There was always a unique cheapest parking lot present on the map. The cheapest rate varied randomly between 0.50 and 1.20 EUR. Increments were 0.10 or 0.20 EUR. 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 (see Fig. 1).

Eye movements were recorded monocular with a head mounted iView X HED system by SMI (Senso Motoric Instruments) with a sample rate of 50 Hz (see Fig. 3). The system recorded eye movements and a video of the scenery from the subject’s perspective to create a video with overlaid eye movements.

The eye tracking videos (test subjects’ visual field with superimposed eye movements, one video per condition) were analyzed offline by a trained rater who classified the current gazes as being located on the mobile display, on the map, or elsewhere, and determined the direction and duration of every gaze shift. Note that we will use the term gaze very broadly in the following to refer to eye movements that stay focused on the display, although the period might actually consist of several fixations and small saccades across this display. The sequence “eyes switch from display to background map” followed by “eyes switch back to display” was coded as one gaze shift. All single gaze shift events during a trial were aggregated into one value.

For evaluating the usability of the system in general, participants had to complete a questionnaire at the end of the test. The aim was to find out how the participants perceived, the efficiency, learnability, and other aspects of the system. The questionnaire was based on the Software Usability Measurement Inventory (SUMI) [24] with a three-point rating scale. The 25 most suitable out of 50 items of the SUMI were selected and adapted to the experimental context. Items were, for example: “I would recommend the phone lens to my colleagues” (scale affect); “The phone lens reacts too slowly” (scale efficiency); “Warnings and error messages are not sufficient” (scale helpfulness/global); “I often need help when using the phone lens” (scale learnability); and “It is easy to use the phone lens for what I want” (scale control). Possible answers were: “I agree,” “I do not know,” or “I do not agree.”

4.4 Participants

The study was conducted with 16 participants (8 female, 8 male). They were students recruited at a local university with a mean age of 26.4 years (SD 2.5 years). None of the participants was familiar with the city, the map, or the application.

4.5 Procedure

Initially, participants were given a short written description of the experiment and the instruction for the subsequent task. Next, the height of the map and pattern sheets was adjusted such that their center was about at shoulder height. After a short practice trial phase for navigating with the mobile phone for each background the eye tracking device was calibrated for the respective participant. This initial phase took about 15 min. Then, eye movement recording was started by the experimenter and the actual test began. Participants had to complete ten trials per condition. After each trial, there was a pause screen that informed the participants about the success of the previous trial and the number of completed trials in this block. When 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 each block, the experimenter asked the participants how they managed the use of the device and how they liked navigation.

After completing the actual test with all four conditions, participants were asked to complete a questionnaire

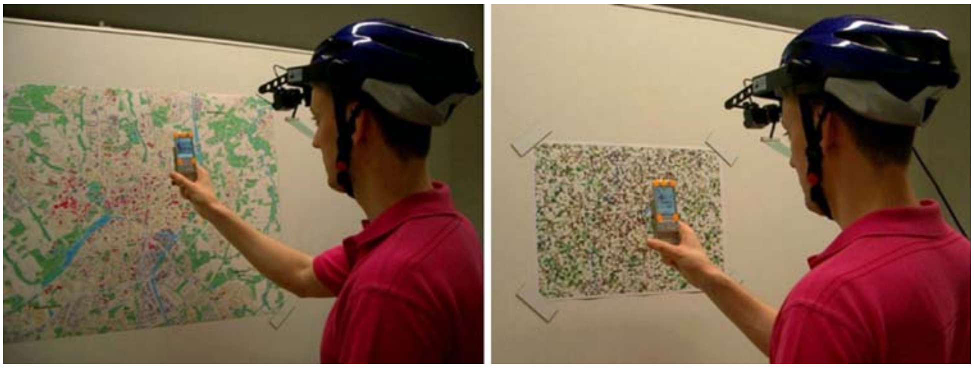

Fig. 3 Participant with eye tracker searching on the large map (*left*) and the small pattern (*right*)

comparing all conditions. It contained a rating in German school grades (“What school grade would you give to this condition?”: 1 = very good, 6 = not sufficient) and open questions asking what they liked and disliked about the navigation for each condition. Finally, participants had to fill in the questionnaire for evaluating the usability of the application.

4.6 Results

All participants completed the experiment. Trial times, error rates, and attention shifts were the main performance measures taken. In addition, motion traces were captured and analyzed to investigate the strategies participants used to explore the search space.

4.6.1 Search time and errors

A histogram of search time suggested that the data were log-normally distributed. Hence all means, confidence intervals, and ANOVAs were computed on the log-transformed data. For the sake of clarity, the descriptive values and graphs are based on retransformed log values. Outliers of more than three SDs from the mean were excluded. Nine of altogether 634 trials were removed this way. Analyses of variance were computed using the “mixed model” function in SPSS.

The average search time over all conditions, measured from the start of a trial until a selection was made, is 40.4 s (95% confidence interval: 39.0–41.9 s). If the user did not select the cheapest parking lot in a trial, then this was counted as an error. The overall error rate is 17.9% (95% confidence interval: 14.9–20.9%). Grouping the results by background type and size shows a strong effect: The large pattern shows the longest search time (64.6 s), followed by large map (47.5 s), small pattern (36.2 s), and small map (31.7 s). A two-factor within-subject ANOVA shows a main effect of background size ($F_{1,246} = 131.73$, $p < 0.01$), a main effect of background type ($F_{1,116} = 21.60$, $p < 0.01$) and an interaction effect of size with type ($F_{1,190} = 5.56$, $p = 0.019$). Merging context and size into a four-stage factor “background” ($F_{3,174} = 44.95$, $p < 0.01$) corresponding to the categories in Fig. 4, pairwise comparisons reveals that the difference between A1 pattern and A1 map is significant ($p = 0.027$), whereas the difference between A3 map and A3 pattern is not ($p = 0.414$, Sidak-adjustment for multiple comparisons).

Providing visual context on the large background leads to a 26% reduction in trial time, whereas visual context for the small size only leads to a reduction by 12%. The item density for the large background is 158.7 items per square meter. For the small background it is 40.4 items per square meter. In the second study the item density was systematically varied to chart the parameter space.

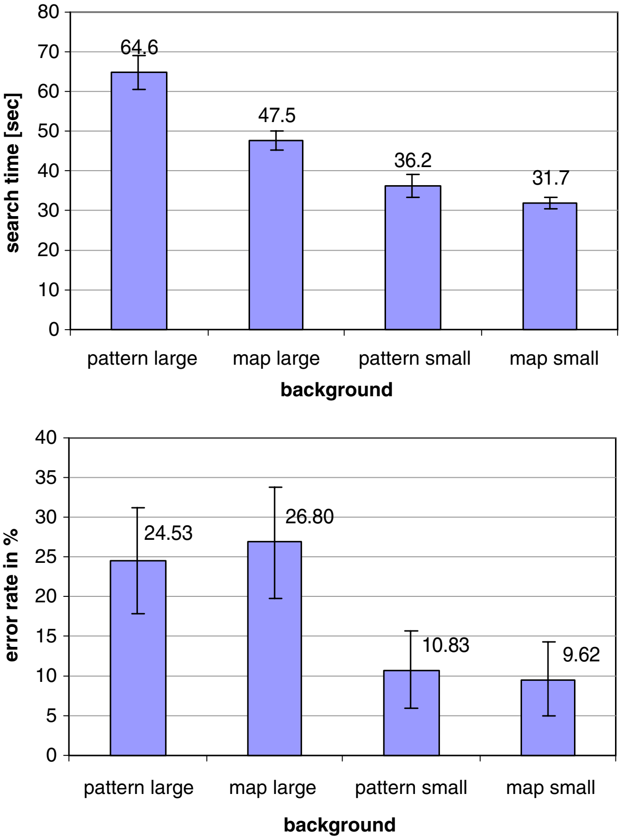

Fig. 4 Search time per trial for each condition (*top*). Error rate per condition (*bottom*)

The differences in error rate (see Fig. 4, right) are also significant. A two-factor within-subject ANOVA asserts a main effect of size on error rate ($F_{1,239} = 20.96$, $p < 0.01$), but no main effect for background ($F_{1,197} = 0.12$, $p = 0.914$) and no interaction effect ($F_{1,219} = 0.207$, $p = 0.650$).

4.6.2 Motion traces

In order to evaluate the effect of the availability of visual context on search strategies, we investigated the motion trajectories on the map. The conditions without visual context typically lead to uniform exploration of the search space at more or less constant speed. Users tend to systematically move over the map in horizontal or vertical zig–zag fashion.

When visual context is available, the motion strategy changes in that the subjects cover the area between items very quickly and spend most of the time inspecting the items. With visual context, the exploration of the search space is more strongly guided by the positions of the items.

4.6.3 Gaze shifts (video coding)

Figure 5, left, shows that the number of gaze shifts drastically increases when background information is provided that can be incorporated in the search ($F_{3,36} = 73.65$, $p < 0.01$). Large and small patterns are both only rarely paid attention to (1–2 times per minute, $p = 0.997$ in Sidak pair-wise comparison). The small map however is already fixated 17.5 times per minute ($p < 0.01$ compared to the pattern conditions), and this number again significantly increases for the large map (37.0 times per minute, $p < 0.01$ to all other conditions in Tukey pairwise comparison).

The two pictures in Fig. 6 serve to illustrate the sequence typically observed with the maps as background: after the price of current symbol is checked, the eyes look ahead to the next possible item (Fig. 6, left) and the hand is moved accordingly until the new item is visible on the mobile display to check its price (Fig. 6, right).

With increasing map size, the eyes need to scan the background longer to find the next possible item. Figure 5, bottom, shows the proportion of background viewing time for all conditions. The overall effect is $F_{3,41} = 54.48$, $p < 0.01$. Again, pairwise comparison (Sidak) yielded significant differences ($p < 0.01$) between all background versions except for the two pattern conditions. Moreover, scanning time for a single gaze shift increases for the large map compared to the small one. This becomes clear when contrasting the relation of gaze shifts per minute (37.0/17.5 = 2.1, see Fig. 5) with the percentage of time spent on the background for both conditions (16.6/6.1 = 2.7, see Fig. 5, bottom). From the small to the large map the number of gaze shifts per minute increases by a factor of 2.1, whereas the time spent viewing the background increases by a larger factor of 2.7. Equivalently, a gaze shift on the large map takes 16.6/37.0 = 45% of the viewing time, whereas a gaze shift on the small map takes only 6.1/17.5 = 35% of the viewing time. The scanning time within one gaze shift is longer and more gaze shifts are performed on the large map.

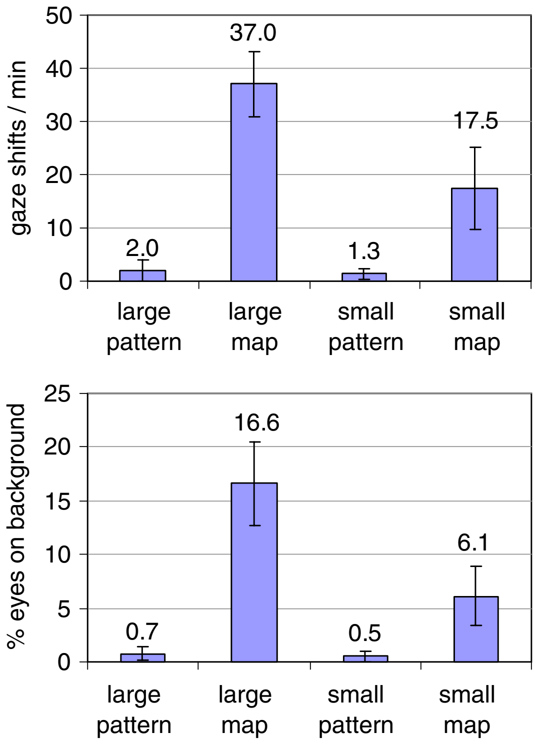

Fig. 5 Gaze shifts per minute per condition (*top*). Gaze on background relative to trial time (*bottom*)

Fig. 6 Gaze—marked by the red eye tracking cursor—on the map (*top*) and on the handheld display (*bottom*). The figure shows two phases of movement towards the target. Gaze shift precedes hand movement (*top*) and hand follows to new gaze position (*bottom*)

4.6.4 Subjective results

After completing all conditions, participants rated them in school grades. A grade of 1 represents the best grade and a grade of 6 the worst. All participants were familiar with this system of school grades.

The order of preference corresponds to the search time results (Fig. 7): Best values were obtained for the small map and worst for the large pattern. According to a Wilcoxon signed rank test, only the difference between small pattern and large map is not significant. Although there is not a large difference between the small pattern and the small map in terms of search time, the subjective rating shows a significant difference between these two conditions ($Z = -2.16$, $p = 0.03$). This means that even though for the small size there was not a significant performance difference, the presence of visual context is still preferred by the participants. For the small map, people liked the possibility of finding the targets quickly because of their spatial proximity and they also liked the need of covering only short distances in terms of motor activity. In contrast, for the large map, the search space was rated as too big for a good orientation and distances were rated as too long. The pattern was not liked because it did not help in orientation at all.

The overall rating for the system is depicted in Fig. 8. It shows that the participants' rating for liking the application (affect) was about average. Efficiency and control over the functions were rated slightly below average. There might be two reasons for this: first, there is a perceptible delay of the tracking system. The majority of the participants rated the application as reacting too slowly. Second, some dexterity and effort are needed for focusing a specific point. Participants stated that keeping the right distance from the surface was sometimes difficult. However, participants were quite satisfied with the clues given by the lens application to help using it (helpfulness) and also with the learnability of the application. The global rating of the application in general was slightly above average, probably negatively influenced by the delay induced by the current implementation and the effort needed to focus the lens.

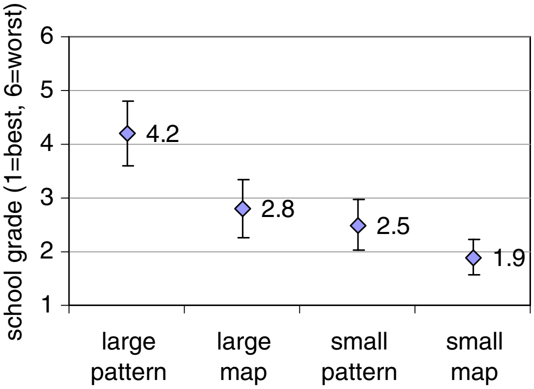

Fig. 7 Participants' ratings per condition (*1* best, *6* worst)

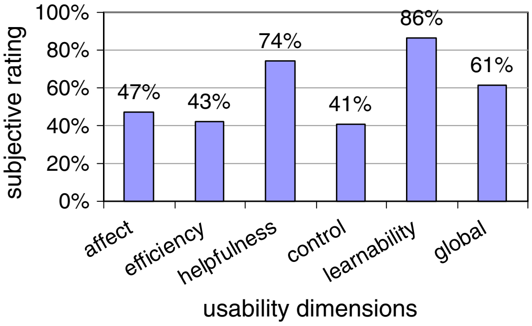

Fig. 8 Participants' rating for the system in general

5 Experiment 2

The second study investigated the effects of the presence of visual context and in particular focused on context size and item density. We again measured completion time, error rate, and satisfaction. The task was the same as in experiment 1. Unlike in experiment 1, the background display was dynamic in that the arrangement of items changed after each trial.

5.1 Design

The study was set up as a 2 × 2 × 4 within-participants design with the following factors (Fig. 9):

1. Visual context: (1) available (map) or (2) not available (pattern).
2. Context size: (1) large (full area) or (2) small (half area).
3. Item count: 2, 4, 8, or 16 items.

5.2 Tasks

The task was the same as in the first experiment, but the positions of the items were randomly changed after each

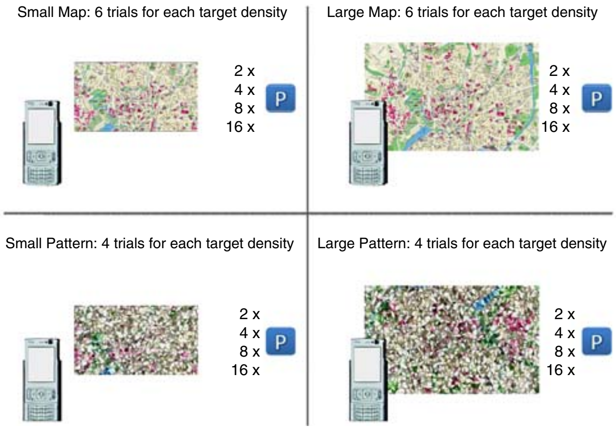

Fig. 9 Experimental conditions of experiment 2

trial. A single trial consisted of scanning the map in the defined condition and finally selecting the target. After finishing six trials per item density in the magic lens condition or four trials per item density in the condition with no visual background, a screen informed the participants about the next condition. The first experiment revealed that the search strategies for the pattern conditions are quite uniform and completion time is much higher than in the map condition. Hence, in order to keep the time for participants in a sensible range, we decided to use only four trials per pattern condition leading to an asymmetric design.

5.3 Apparatus

The experiment was performed on a Nokia N95-8GB camera phone. The client application showed the augmented view of the map and captured all user interactions and movements with timestamps. The background was displayed on a Barco LCN-42 LCD screen (42″, 1,920 × 1,080 pixels, 93 × 53 cm). For the condition without visual context the phone was tracked above an abstract colored pattern (Fig. 9, bottom). For the condition with visual context a colored city map was shown (Fig. 9, top). For the large size the background filled the whole area (0.492 m^2) of the 42″ display (Fig. 9, right). For the small size half of the display area was used (0.246 m^2) (Fig. 9, left).

Eye movements were recorded binocular with a head mounted Eyelink 2 system by SR Research with a sample rate of 250 Hz. As in the previous study, eye movements and a video of the scenery from the subject's perspective were recorded simultaneously to create a video with overlaid eye movements. In addition, the raw signal of the eye movement was processed to identify fixations and saccades (gaze jumps). The increased spatial and temporal resolution compared to the first study allowed for a detailed analysis of saccadic amplitudes and fixation parameters.

5.4 Participants and procedure

The study was conducted with 17 participants, 12 female, 5 male. They were students recruited at a local university, aged 20–31 years (mean age 26.4 years). None of the participants had taken part in the first experiment. No participant was familiar with the city map.

Initially, participants were given a short written description of the experiment and the instruction for the subsequent task. After that, a 5–7 min practice period for navigating with the mobile phone for each condition followed. Participants had to complete the requested number of trials per condition. As in the first experiment, participants were asked for feedback after each block and filled out a final questionnaire after having completed all conditions.

5.5 Results

All participants were able to perform the experimental tasks. As before, trial times, error rates, and attention shifts were the main performance measures taken.

5.5.1 Search time and errors

A histogram of trial time suggested a log-normal distribution and hence the analysis operates on the log-transformed data. For the sake of clarity, the graphs show the retransformed means. Outliers of more than three SDs from the mean were excluded. Eleven outlier trials were removed in this way.

The overall time per trial, measured from the start of a trial until a selection was made, was 26.7 s (95% confidence interval: 26.2–28.6 s). If the user did not select the cheapest parking lot in a trial, then this was counted as an error. The overall error rate was 12.4% (95% confidence interval: 10.5–14.2%). A three-factor within-subject repeated-measures ANOVA shows main effects on search time for all factors (availability of visual context: $F_{1,216} = 141.2$, $p < 0.001$; background size: $F_{1,178} = 56.4$, $p < 0.001$; item count: $F_{3,367} = 59.9$, $p < 0.001$).

Figure 10a, b shows the average trial times and error rates by background type. The small map ("m") takes the least amount of time (17.1 s), followed by the large map ("M") with 23.3 s. The small ("p") and large ("P") pattern take 30.9 s and 52.7 s, respectively. These times are pairwise significantly different (Sidak-adjustment for multiple pairwise comparisons). The error rates for the small map (8%) and pattern (9%) are comparable, those for the large map and the large pattern increase to 13 and 19%, respectively. Providing visual context for the small background thus reduces the search time by 44.4%. For the large background, the reduction is 44.2%. The reduction in error rate is 13.5% for the small and 31.2% for the large background, respectively.

Figure 10c shows that the search time increases with the number of items. There is an interaction effect between the availability of visual context and the number of items ($F_{3,342} = 5.2$, $p = 0.002$). This suggests that the slope of search time with increasing item count depends the availability of visual context. Figure 10d shows the search time per item count broken down by background type. As expected, the large pattern ("P") takes longest, followed by the small pattern ("p"), the large map ("M"), and the small map ("m").

The average density of items on the display (number of items divided by display size) was thus for the large size 4.1, 8.1, 16.3, and 32.5 items per square meter and for the small size 8.1, 16.3, 32.5, and 65.0 items per square meter.

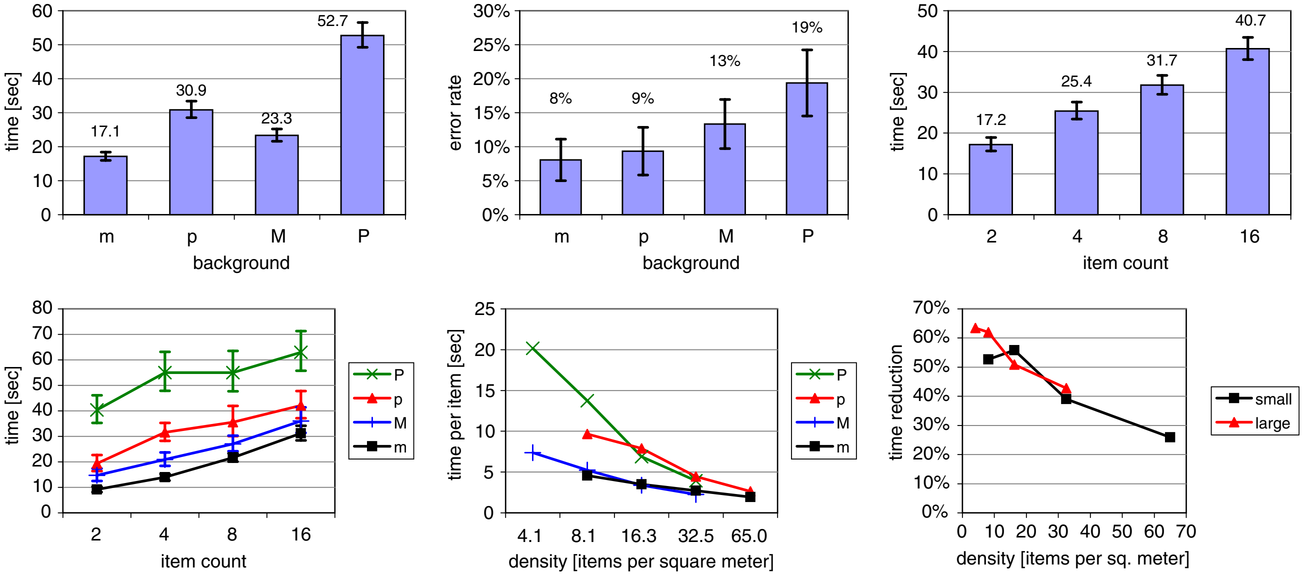

Fig. 10 Results: **a**, **b** time and error rate by background, **c**, **d** time by item count, **e** time by density, **f** time reduction by density. *m* small map, *M* large map, *p* small pattern, *P* large pattern

Figure 10e shows that the search time per item decreases with increasing item density. This is as expected, because the higher the density, the smaller the area the user has to scan in order to find the next item. Interestingly, irrespective of the background size, for the overlapping item densities (8.1, 16.3, 32.5) the search times per item for the conditions with visual feedback (magic lens, “m”, “M”) are very close. This is also the case for conditions without visual feedback (dynamic peephole, “p”, “P”). Moreover, the overall times for the dynamic peephole are higher than for the magic lens, but their performances converge as density increases.

For the highest density (65.0) there is only a small advantage of the magic lens over the dynamic peephole. This result is in line with experiment 1, which did not find a significant difference in search times for a background of size A3 and a density of 158.7 items per square meter.

The interaction between density and visual context means that the time reduction that can be expected from using visual context (i.e. using a magic lens interface rather than a dynamic peephole interface) decreases as density increases.

5.5.2 *Gaze shifts (video coding)*

Corresponding to experiment 1, the eye tracking videos were manually coded by a trained rater to distinguish between “gaze on mobile phone display” and “gaze on background”. The results of the video coding are shown in Fig. 11: Almost no gaze shifts occur with the large or small pattern as background. For large or small map backgrounds, the proportion of time the gaze is on the background decreases with item count (Fig. 11, top) or item density, respectively (Fig. 11, bottom).

An ANOVA (three-factor within-subject repeated-measures) shows main effects for availability of visual context ($F_{1,220} = 386.3$, $p < 0.001$), item density ($F_{4,646} = 4.06$, $p = 0.003$), and the interaction of both factors ($F_{4,491} = 3.8$, $p = 0.005$), but not for background size ($F_{1,140} = 0.393$, $p = 0.532$) or other possible interactions.

5.5.3 *Eye movements*

In addition to the video with overlaid gaze location, the higher temporal resolution of the eye tracker used in the second experiment also allowed for a more detailed analysis of oculomotor parameters like saccades (gaze jumps) and fixations. These events were automatically identified in the raw signal using the default setting of the Eyelink 2 system. Saccades were divided in two groups based on their amplitude: Any saccade larger than 7° (corresponding to the diagonal of the mobile display) was assumed to involve some background scanning (either from the mobile to the background, on the background itself, or back from the background to the phone display), whereas saccades smaller than that were classified as being on the phone. This criterion is conservative with regard to background usage as saccades smaller than 7° on the background are treated as being on the phone.

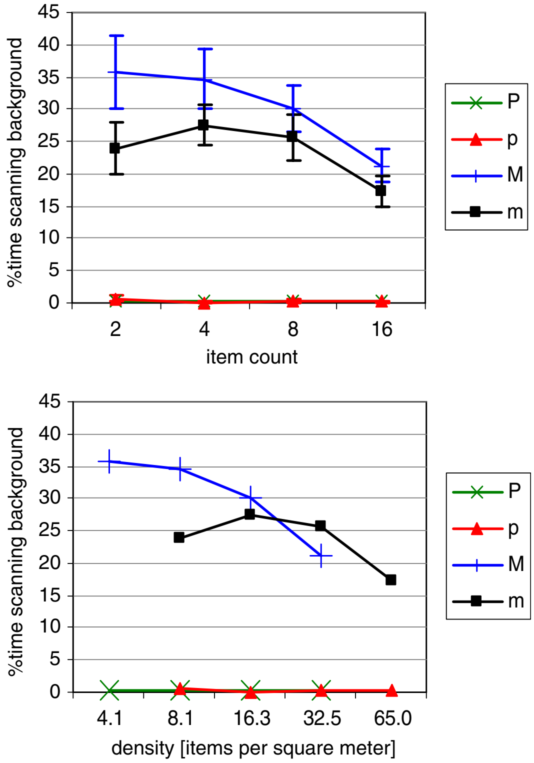

Fig. 11 Results of coded eye tracking videos. Proportion of time the gaze is scanning the background. *Left* Item count, *right* item density. *Lines* represent different backgrounds. *m* small map, *M* large map, *p* small pattern, *P* large pattern

Still, the splitting reveals the same trend apparent in the manual video analysis (Fig. 12) with increasing item density, the percentage of saccades involving background scanning decreases for the large and small map.

An ANOVA (three-factor within-subject repeated-measures) shows effects for availability of visual context ($F_{1,227} = 459.5$, $p < 0.001$), item density ($F_{4,687} = 9.01$, $p < 0.001$), and the interaction of both factors ($F_{4,521} = 3.5$, $p = 0.008$), but not for background size ($F_{1,135} = 0.340$, $p = 0.561$) or other possible interactions. Both parameters—percentage of time the gaze is directed to the background (video coding) and percentage of saccades >7° within a trial—correlate with $r = 0.811$ ($p < 0.001$).

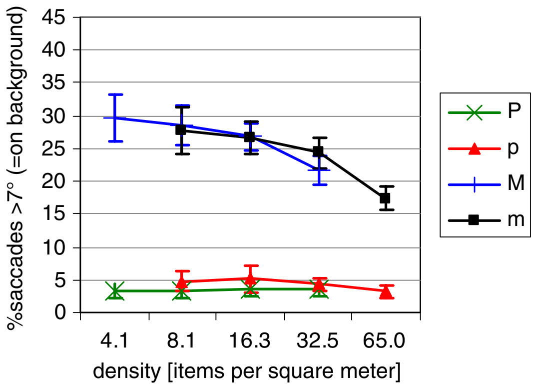

Fig. 12 Proportion of saccades >7° (classified as "on the background") over item density. *m* small map, *M* large map, *p* small pattern, *P* large pattern

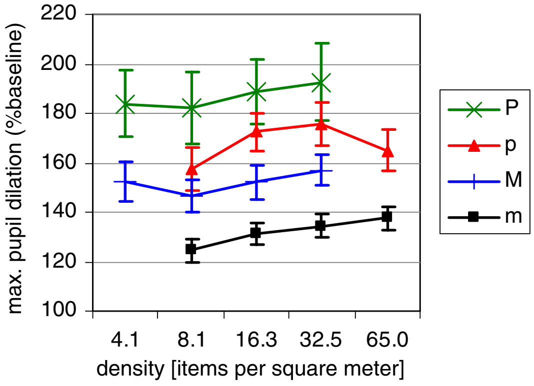

Fig. 13 Maximum pupil dilation during a trial over item density expressed as percent baseline dilation (dilation during first second of stimulus onset per background type). *Lines* represent different backgrounds. *m* small map, *M* large map, *p* small pattern, *P* large pattern

The maximum pupil dilation during a trial was used to assess mental workload during the search task [25]. As pupil dilation is influenced by stimulus luminance and our stimulus material varied slightly in luminance (mean values determined with Adobe Photoshop: P: 199, p: 205, m: 221, M: 224), pupil dilation during the first second after stimulus onset in each condition (background type) served as a baseline value for each background type. Figure 13 shows an increase in maximum pupil dilation for each background type ranging from small map to large pattern.

An ANOVA (three-factor within-subject repeated-measures) confirms all main effects for availability of visual context ($F_{1,249} = 102.9$, $p < 0.001$), item density ($F_{4,730} = 9.2$, $p < 0.001$), and background size ($F_{1,89} = 13.2$, $p < 0.001$), but no interactions.

5.6 Subjective results

Subjective feedback was collected for the second experiment as well. The same questionnaire based on the SUMI was used. The results are very similar to the ones described for experiment 1 (best grade for small map, worst grade for large pattern) and show no new trends. For this reason this is not discussed further here.

6 Discussion

The performance results of the first experiment, namely search time and error rate, suggest that for small backgrounds and high item densities there is no substantial advantage of "meaningful" visual context beyond the device display over an abstract background pattern for the visual search task we investigated. This may be partly due to the relatively large portion of the background that is occluded by the device, which leaves less visual context in comparison to large backgrounds. As the second experiment with lower target densities shows, this result is also due to the eye movement strategy used by most participants for the small map size and the large number of items. Subjects rarely looked at the background but predominantly stayed focused on the device display. In the experimental task, often more than one item was visible on the mobile display, so there was little need to scan the background map for the next item and the advantages of having a constant fixation object (the mobile display) that was guided slowly from item to item apparently prevailed. For the large backgrounds in experiment 1, the context helps in navigation time, yet the effect is less clear than presumed and may be influenced by the nature of the task. The search task was chosen to represent one of the most basic interactions with a map: finding a target randomly positioned on the map. The more complex the search becomes (i.e. finding a restaurant with view on the river), the more benefit can be expected from incorporating additional structural information provided by the map.

This search strategy is strongly influenced by the availability of visual context, which is reflected in the motion traces we recorded: while on the abstract pattern subjects moved the phone systematically to cover the whole area (resembling “window cleaning” movements) and adjusted their movements to obtain regular display updates, the traces in which the items were visible on the background appear much more “jumpy.”

The second experiment shed light on the role of item density on the user’s ability to take advantage of the information present on the background display. For lower target densities the background is used more intensely, the number of gaze shifts and the time spent with gaze on the background increases. The results of the second experiment showed that the amount of time the gaze is located on the background depends first on the availability of meaningful information (pattern vs. map) and then on the density of relevant items, but not on the size of the background.

The results of manual video coding (Fig. 11) and automatic classification of saccades in “on the display” (≤7°) or “involving background scanning” (>7°) (Fig. 12) resemble each other strongly and both indicate that for searching items, visual scanning is immediately utilized if possible. Thus an interface may benefit from supporting this interaction modality.

Figure 13 shows an increase in maximum pupil dilation for each background type ranging from small map to large pattern that corresponds to subjective evaluation. Within one background type, maximum pupil dilation during a trial rises with item density. Note that this cannot be attributed to a mere time-on-task-effect, as the sequence of item density was varied across subjects. Porter et al. [25] attribute the increase observed in their study to spatial memory demands. This interpretation might also have implications for future applications, which we will address in our conclusion.

7 Conclusion

This paper presented a study on the effects of visual context for magic lens and dynamic peephole interactions in a basic object locator task. The main factors tested were the availability of visual context, the size of the context, and the number of items users had to investigate. In the case with visual context the items were visible on the background surface, but the attribute to look for was only available via the magic lens. Users in this case had the option to scan for the items on the background or to use the magic lens. For deciding whether the item was the right one, they had to inspect it with the mobile device.

We found that the effectiveness of visual context does not primarily depend on its physical size, but on the density of the items. Visual context is most effective for sparsely distributed items. The denser the items are distributed, the less clear the performance benefit that can be expected from providing visual context. One reason for this seems to be that for high densities it is more likely that the next item already appears at the border of the display, hence making a switch of visual attention to the background unnecessary.

High item densities also result in lower average distances between the items. Thus the next item may already appear in the visual periphery of the user, even though it is not yet located on the device display. In such cases there seems to be a tradeoff between shifting one’s gaze from the device display to the background and moving the hand. For relatively close items moving the hand towards the item in the visual periphery may be the more efficient strategy, compared to switching visual attention from the device display to the background. Switching attention for visual search on the background incurs some cost, because of the need to refocus on the new layer of presentation at another distance. An important related result is a study which revealed that people have difficulties to disengage attention from objects that are near their hands [21]. Further research is needed to clarify this tradeoff.

Both subjective as well as behavioral data (search time and error rate) show that mobile navigation interfaces benefit from a magic lens option to interact with public maps, especially if theses maps are supposed to cover a large area or items are distributed sparsely.

Given these results it is advisable to constrain the item density in magic lens interfaces if the designer wants to ensure that users also pay attention to the background itself (e.g. in the case of advertisements). This can be achieved, for example, by performing suitable pre-filtering of information categories to limit the number of candidate items. When the item density is too high, no significant performance benefits can be expected from external visual context, although visual context is preferred in this case as well. Wickens et al. [26] give additional advice how to improve map design for visual search in the case of cluttered displays. Future users on the other hand may be advised that from a certain density switching between background and mobile phone display may lead to worse search performance. They should rather try just to remain focused on the mobile phone display.

The larger the number of items becomes, the higher is the cognitive effort for memorizing all visited items and the current "best" (according to the search task) item. An application could support the user by implementing a memory aid that highlights the already visited items or that allows the user to mark particular items. Moreover, users could be guided towards items on the map using halos or similar techniques.

For practical applications where user acceptance is crucial, one should keep in mind that the subjective results show a clear preference for the real map compared to a pattern without visual context of the same size (see Fig. 7), probably because visual scanning is still the most natural way to search. Any new service should pick up on existing preferences instead of enforcing new and unfamiliar strategies like the "window cleaning" pattern we observed during search on the abstract pattern. Moreover, when augmenting already available public maps that are used as part of leisure activities, no one wants to educate potential users to establish systematic search behavior. Instead it is preferable to enable active and spontaneous interaction with an additional layer of information as is possible with the investigated magic lens interface.

Acknowledgments We would like to thank Sandra Trösterer from the Center of Human–Machine–Systems, Berlin, for helping us running the first experiment and Ahmad Abbas and Robert Walter (both from the Deutsche Telekom Laboratories, Berlin) for their assistance in data recording and processing.

References

1. Reilly D, Welsman-Dinelle M, Bate C, Inkpen K (2005) Just point and click?: using handhelds to interact with paper maps. In: Proceedings of MobileHCI '05. ACM Press, New York, pp 239–242
2. Schöning J, Krüger A, Müller HJ (2006) Interaction of mobile camera devices with physical maps. In: Adjunct proceedings of Pervasive'06. Austrian Computer Society (OCG), Vienna
3. Rohs M, Schöning J, Krüger A, Hecht B (2007) Towards real-time markerless tracking of magic lenses on paper maps. In: Adjunct proceedings of Pervasive '07. Austrian Computer Society (OCG), Vienna
4. Bier EA, Stone MC, Pier K, Buxton W, DeRose TD (1993) Toolglass and magic lenses: the see-through interface. In: Proceedings of SIGGRAPH '93. ACM Press, New York, pp 73–80
5. Rohs M, Schöning J, Raubal M, Essl G, Krüger A (2007) Map navigation with mobile devices: virtual versus physical movement with and without visual context. In: Proceedings of ICMI '07. ACM Press, New York
6. Mehra S, Werkhoven P, Worring M (2006) Navigating on handheld displays: dynamic versus static peephole navigation. ACM Trans Comput Hum Interact 13(4):448–457
7. Rekimoto J (1995) The magnifying glass approach to augmented reality systems. In: Proceedings of ICAT/VRST '95, pp 123–132
8. Yee KP (2003) Peephole displays: pen interaction on spatially aware handheld computers. In: Proceedings of SIGCHI '03. ACM Press, New York, pp 1–8
9. Fitzmaurice GW (1993) Situated information spaces and spatially aware palmtop computers. Commun ACM 36(7):39–49
10. Hachet M, Pouderoux J, Guitton P, Gonzato JP (2005) TangiMap: a tangible interface for visualization of large documents on handheld computers. In: Proceedings of GI '05, pp 9–15. Canadian Human-Computer Communications Society
11. Baudisch P, Good N, Bellotti V, Schraedley P (2002) Keeping things in context: a comparative evaluation of focus plus context screens, overviews, and zooming. In: Proceedings of CHI '02. ACM Press, New York, pp 259–266
12. Sanneblad J, Holmquist LE (2006) 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. ACM Press, New York, pp 373–377. http://doi.acm.org/10.1145/1133265.1133343
13. Hecht B, Rohs M, Schöning J, Krüger A (2007) Wikeye—using magic lenses to explore georeferenced Wikipedia content. In: Proceedings of PERMID '07
14. Ellison A (2008) Are results from different techniques mutually exclusive in the study of how the brain processes visual search? Cortex 44(1):99–101
15. Thornton TL, Gilden DL (2007) Parallel and serial processes in visual search. Psychol Rev 114(1):71–103
16. Treisman AM, Gelade G (1980) A feature integration theory of attention. Cogn Psychol 12:97–136
17. Wolfe JM (1994) Guided search 2.0—a revised model of visual search. Psychon Bull Rev 1(2):202–238
18. Jaschinski W, Heuer H, Kylian H (1998) Preferred position of visual displays relative to the eyes: a field study of visual strain and individual differences. Ergonomics 41(7):1034–1049
19. Findlay JM, Brown V, Gilchrist ID (2001) Saccade target selection in visual search: the effect of information from the previous fixation. Vis Res 41:87–95
20. Araujo C, Kowler E, Pavel M (2001) Eye movements during visual search: the costs of choosing the optimal path. Vis Res 41(25–26):3613–25
21. Abrams RA, Davoli CC, Du F, Knapp WH, r, Paull D (2008) Altered vision near the hands. Cognition 107(3):1035–47
22. Van Orden KF, Limbert W, Makeig S, Jung TP (2001) Eye activity correlates of workload during a visuospatial memory task. Hum Factors 43(1):111–121
23. Reichenbacher T (2001) Adaptive methods for mobile cartography. J Geogr Sci 11(Suppl 2001):43–55

24. Kirakowski J, Corbett M (1993) SUMI: the software usability measurement inventory. Br J Educ Technol 24(3):210–212
25. Porter G, Troscianko T, Gilchrist ID (2007) Effort during visual search and counting: insights from pupillometry. Q J Exp Psychol 60(2):211–229
26. Wickens CD, Alexander AL, Ambinder MS, Martens M (2004) The role of highlighting in visual search through maps. Spat Vis 17(4–5):373–388