Investigating Gaze-supported Multimodal Pan and Zoom

Sophie Stellmach* and Raimund Dachselt†
User Interface & Software Engineering Group
Faculty of Computer Science
University of Magdeburg, Germany

Abstract

Remote pan-and-zoom control for the exploration of large information spaces is of interest for various application areas, such as browsing through medical data in sterile environments or investigating geographic information systems on a distant display. In this context, considering a user's visual attention for pan-and-zoom operations could be of interest. In this paper, we investigate the potential of gaze-supported panning in combination with different zooming modalities: (1) a mouse scroll wheel, (2) tilting a handheld device, and (3) touch gestures on a smartphone. Thereby, it is possible to zoom in at a location a user currently looks at (i.e., gaze-directed pivot zoom). These techniques have been tested with Google Earth by ten participants in a user study. While participants were fastest with the already familiar mouse-only base condition, the user feedback indicates a particularly high potential of the gaze-supported pivot zooming in combination with a scroll wheel or touch gesture.

CR Categories: H.5.2 [Information Interfaces and Presentation]: User Interfaces - Evaluation/methodology—Input devices and strategies;

Keywords: Eye tracking, multimodal, gaze-supported interaction, touch, tilt

1 Introduction

Panning and zooming are essential for the exploration of diverse information spaces, such as large images and geographical information systems (GIS). In this context, our gaze is ideal to indicate a user's current point-of-interest (PoR), for example for specifying where to zoom in (e.g., [Hansen et al. 2008]). However, several challenges are associated with gaze interaction, such as involuntarily performing an action (i.e., *Midas Touch* problem) and imprecise gaze data. These can be overcome with multimodal gaze-supported interaction (e.g., [Castellina and Corno 2008; Stellmach and Dachselt 2012]). However, despite this high potential, thorough investigations on suitable input and output combinations benefitting from gaze as a supporting modality are still insufficient.

Motivated by this, we investigate how gaze input can be combined well with a smartphone to remotely pan and zoom on a distant display. Smartphones are of special interest as they are easy to use, commonly available, and can be used in various application contexts. In addition, as the mouse is still prevalent for the interaction with desktop computers, we also want to find out how gaze could be used in combination with well-established scroll-based input. For this purpose, we describe and compare five different pan-and-zoom techniques (four of them are gaze-supported) using the example of Google Earth. While several studies have investigated gaze-based pan-and-zoom techniques, only few have examined gaze in combination with handheld control devices. We contribute to a better understanding of gaze-supported interaction. This is especially interesting for contexts in which traditional mouse and keyboard input may not be available or even feasible, as for the interaction with public displays, large-sized TV sets or see-through glasses.

*e-mail:stellmach@acm.org
†e-mail:dachselt@acm.org

2 Related Work

Several works deal with gaze-based pan and zoom, such as Lankford [2000] who proposes a magnification tool based on gaze dwell-time. Hansen et al. [2008] and Adams et al. [2008] present a gaze-controlled pan and zoom interface for which they use a discrete central zoom region surrounded by a pan region towards the screen border. Zhu et al. [2011] take advantage of the entire screen space for panning for which the panning speed depends on the distance between screen center and current PoR.

Further approaches exist for eye-controlled zooming interfaces with an additional manual activation. For example, Bates and Istance [2002] propose eye-controlled zooming user interfaces (UI) to facilitate accessing mainstream graphical UIs for which the entire screen is magnified and the zoom is manually controlled. In addition, Fono and Vertegaal [2005] present EyeWindows for which a currently fixated window increases in size (zoomed in) and peripheral windows are miniaturized (zoomed out). For the zoom activation users preferred pressing a keyboard button over gaze dwelling. Adams et al. [2008] compare four different pan and zoom input techniques including gaze-based panning with zooming via clicking a certain mouse button, moving the head towards or away from the screen, and a gaze dwell-based activation. While none of the gaze-based methods proved to be as efficient as the conventional mouse-based input, user feedback for these techniques was encouraging. However, a critical discussion of their proposed techniques is lacking and thus it is unclear how they would benefit user contexts in which mouse input may not be available.

In [Stellmach et al. 2011], we propose a combination of gaze and an iPod touch for a gaze-supported exploration of large image collections on a distant display. There we proposed a *Look-Touch-Tilt* zooming, for which the user can directly zoom in at the PoR by activating the zoom mode via a touch and then tilting the iPod. A touch event is used to affirm the intention to zoom. For panning, we proposed a combination of a relative touch-based panning on a mobile screen and looking at the screen borders of a distant display. However, while we provided comprehensive qualitative user feedback on how users conceived these interaction techniques, a thorough empirical evaluation of their proposed techniques is missing.

Finally, Nancel et al. [2011] compare several mid-air pan-and-zoom techniques (without gaze input). The two fastest techniques were a touch-based zoom gesture on a mobile device and a scroll wheel zooming with the secondary hand, while the zoom focus and panning direction were set by pointing with the primary hand.

3 Pan and Zoom Alternatives

Motivated by the promising potential of gaze-supported interaction with a flexible mobile device such as a smartphone, we decided to investigate gaze input in combination with touch gestures on and tilt gestures with a smartphone, and a mouse scroll wheel. This way, we want to find out how users would assess these modalities for gaze-supported zooming, and it offers a high potential for further investigations for application contexts in which mouse input may not be available. In addition, we are interested in how users assess gaze panning compared to mouse- and touch-based panning. For this purpose, we have elaborated and implemented five variants of pan-and-zoom combinations, which are briefly explained in the following and compared in a user study reported in Section 4.

Scroll Wheel Zooming + Mouse Panning (Sc+M). As a base condition we use a mouse to zoom via its scroll wheel and to pan by holding the left mouse button down and dragging the view into the desired direction. Hence, this mimics the conventional Google Earth mouse control. However, it leaves out possibilities such as double clicking to quickly zoom in on a hovered location. If turning the scroll wheel forward, the view will be zoomed in at the current cursor location.

Scroll Wheel Zooming + Gaze-directed Panning (Sc+G). Here the zooming works the same as for condition **Sc+M**, but this time the view is zoomed in towards the current PoR (e.g., similar to [Hansen et al. 2008]). For the gaze-pivot zooming [Stellmach et al. 2011] several aspects have to be taken into account for setting an appropriate rate at which viewed content moves towards the screen center. This means that in contrast to ordinary pivot zooming using a mouse, very quick panning motions should be prevented to reduce disorientation and motion sickness. The panning is directed via gaze, so that currently viewed content moves towards the screen center. The panning speed is dependent on the distance to the screen center (similar to [Zhu et al. 2011]). Finally, to prevent a Midas touch effect, the user can toggle the gaze panning by briefly pressing the scroll wheel. Thus, if this mode is inactive and the scroll wheel is not used, the gaze input will not have any effect.

Touch-based Zooming + Gaze-directed Panning (To+G). For **To+G**, the zooming is controlled via a simple touch gesture on the mobile device. If touching the mobile screen and moving the finger upward/downward, the view will be zoomed in/out. A *relative touch* approach is used as described by Stellmach et al. [2011]. This means that the user can touch anywhere on the mobile screen and can perform the zooming based on this initial touch position. The further the current touch position is away from the initial one, the faster is the zooming. The gaze-directed panning for **To+G** works the same way as for **Sc+G**. To prevent a Midas touch effect, gaze panning is only active while touching the mobile screen.

Tilt-based Zooming + Gaze-directed Panning (Ti+G). For zooming in the **Ti+G** condition, the handheld needs to be tilted forward or backward. Depending on the underlying users interaction metaphor, a tilting forward could be interpreted as zooming in (metaphor: *dive into*) or zooming out (metaphor: *push away*). In addition, a *relative tilt* is used for a higher flexibility and to reduce straining a users wrist. This means that the orientation of the handheld when first touching the mobile screen is used as a null reference. The larger the tilting angle with respect to this initial orientation, the faster is the zooming. Similar to the previously described input conditions, a gaze-pivot zoom is used, so that a user can directly zoom in on an object of interest. The gaze-directed panning for **Ti+G** works the same way as for **Sc+G** and **To+G**. In order to not activate something unintentionally while gesticulating with the handheld device, both zooming (via tilt) and panning (via gaze) require touching the mobile display.

Tilt-based Zooming + Gaze & Touch Panning (Ti+GT). Finally, we wanted to find out whether users would rather use a touch-based panning instead of their gaze if they had the choice. For this purpose, condition **Ti+GT** builds on **Ti+G** and additionally offers the possibility to pan via touch input. Users can simply touch the mobile screen (anywhere) to activate the gaze-directed panning. At the initial touch position no touch panning will be performed and the gaze panning is active. If moving the finger for a minimal threshold on the screen (we set it to 50 pixels), the touch panning is activated and the gaze input is not considered anymore. In addition, it is possible to *nudge* the scene. Thus, the panning movement continues, slowly residing, after lifting the finger from the touch screen. The movement will immediately stop, if briefly tapping the mobile screen again. Furthermore, a faster panning can be achieved by increasing the distance from the initial to the current touch position on the handheld.

4 User Study

Based on the initial motivation in the introduction, we conducted a user study testing the specified mouse-based and four gaze-supported pan-and-zoom techniques. We put a particular interest on the user feedback, as a main interest was to find out how users would enjoy and assess the gaze-supported techniques.

Design. A within-subjects design was used with the five described input conditions **Sc+M**, **Sc+G**, **To+G**, **Ti+G**, and **Ti+GT**. The conditions have been tested in a counterbalanced order based on a *Latin square* design.

Participants. Ten participants (3 female, 7 male) volunteered in the study, aged from 21 to 30 (Mean (M) = 25.9) with normal or corrected-to-normal vision. Five participants have used eye tracking a few times before and all participants are daily computer users. Based on a 5-point Likert scale from 1 - *Do not agree at all* to 5 - *Completely agree*, participants had to rate several statements about their background. Based on this, they indicated that they are all familiar with Google Earth and usually use it with its standard mouse input (M=3.7, Standard Deviation (SD)=1.19). In addition, they specified that they mainly use keyboard and mouse at the computer (M=4.8, SD=0.4), but that they are open-minded to novel input modalities (M=4.7, SD=0.46). They are also familiar with touch input, such as on modern smartphones (M=4.5, SD=1.03).

Apparatus. An Apple iPod Touch (2nd generation) is used as a mobile touch-and-tilt device. The Tobii T60 eye tracker, a table-mounted binocular eye tracker that is combined with a 17 inch TFT display, has been used to gather gaze data. It has a screen resolution of 1280x1024, a 0.5° accuracy, and a sampling rate of 60 Hz. The gaze cursor position is stabilized using *speed reduction* [Zhang et al. 2008]. Based on initial tests before the user study, we use a ratio of 5% of the current with 95% of the previous gaze position. For the interaction with a GIS, we use Google Earth. For this purpose, we implemented a Microsoft Windows Forms tool based on C# that uses the Google Earth plug-in. While it is possible to gradually adapt the pan and zoom speed for the touch- and tilt-based zooming, we kept the maximum speed for all five conditions the same.

Procedure. After welcoming a participant, a brief introduction and a demographic questionnaire (e.g., asking about age and familiarity with eye tracking and Google Earth) was handed out. Participants were seated approximately 60 cm from the eye tracker screen and were instructed to sit fairly still, but without actively restricting their movement. A 9-point eye tracker calibration sequence was performed to adjust it to the respective participant. For each pan-and-zoom condition, the same procedure was followed. First, one pan-and-zoom technique at a time was explained and could be tested by the participants until they felt sufficiently acquainted with

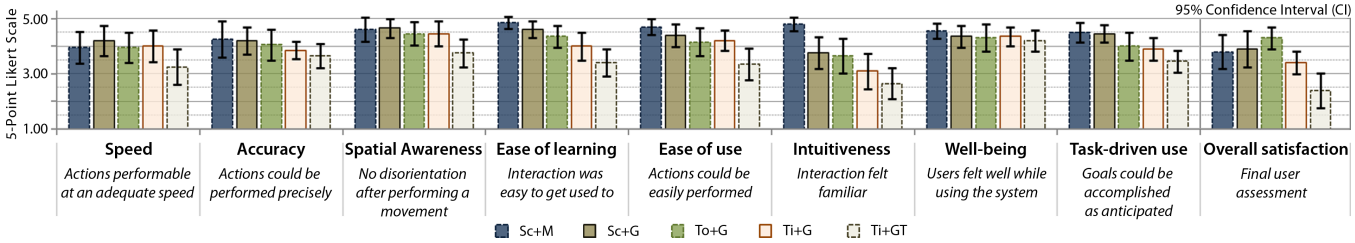

Figure 1: *Qualitative user assessment based on intermediate questionnaires and the overall user satisfaction based on the final questionnaire.*

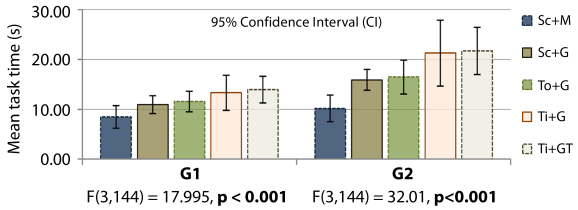

Figure 2: *Task completion times have been summarized into two distinct groups - Group 1 (G1): London, Esher, Buenos Aires, and Carmelo. Group 2 (G2): Tokyo, Niiza, Sydney, and Coogee.*

it (usually less than five minutes). The actual test tasks were to navigate to a certain location which was shown on overview maps on which a city was marked. Eight cities located on four different continents were tested in a random order: London and Esher, Tokyo and Niiza, Sydney and Coogee, and Buenos Aires and Carmelo. A yellow pin in Google Earth marked the respective targets. The default starting position was always the same: 25N 0' 0", 40W 0' 0", and ca. 16 000 miles above the ground which is over the North Atlantic Ocean. Users had to zoom in on a target until a certain height and radius to the target was reached and the program stopped. For this study, we only used such basic search tasks on purpose, because our main focus was not what and when certain zoom and panning actions would be used, but rather how well users would get along with the particular input combinations.

Measures. Our quantitative measures included logged target acquisition times, camera positions, and gaze data. For substantial user feedback, an *intermediate questionnaire* was handed out after the eight cities had been reached with a respective input technique. The *intermediate questionnaires* were the same for all techniques and consisted of two types of questions, for which all quantitative questions were based on 5-point Likert scales from 1 - *Do not agree at all* to 5 - *Completely agree*: (Q1) Sixteen statements had to be rated concerning eight usability aspects (two for each) that are summarized in Figure 1, and (Q2) two questions asking for qualitative feedback on what the users particularly liked and disliked about the tested pan-and-zoom techniques. After having tested all five techniques, participants were asked to assess how they liked the respective modalities for zoom and pan and in the combination of them (again based on the previously described 5-point Likert scale) in a *final questionnaire*. On average, each session took about 80 minutes with instructions, carrying out the described procedure, and completing the questionnaires.

5 Results

For the evaluation of the conducted study we want to point out again that the mouse condition should only be considered as a baseline for comparison. We did not aim at beating it, since all users were already very familiar with this type of input. Instead, we were mainly interested in user feedback that provided great insights into how the interaction felt, how it may be improved, and if these techniques could be a valid alternative to traditional mouse input for contexts in which a mouse may not be suitable.

Task completion times. A repeated-measures ANOVA (Greenhouse-Geisser corrected) and post-hoc sample t-tests (Bonferroni corrected) were used to investigate task completion times. Based on this, we could identify two groups of targets among which target times did not differ significantly (see also Figure 2): (G1) London, Esher, Buenos Aires, and Carmelo and (G2) Tokyo, Niiza, Sydney, and Coogee. For G1, it was often sufficient to perform small panning steps or completely rely on the pivot-based zooming. For G2, longer panning actions had to be performed, which caused longer completion times compared to G1, especially for the gaze-supported techniques. As we have already assumed, users were fastest with the base condition **Sc+M** (see Figure 2). In fact, **Sc+M** was significantly faster than all other techniques for G1 and G2 ($p<0.001$), except for **Sc+G** in G1 ($p=0.015$). The tilt-based conditions **Ti+G** and **Ti+GT** achieved significantly worse results than **Sc+M** ($p<0.001$), **Sc+G** ($p<0.001$), and **To+G** (for G1 $p<0.001$; for G2 $p<0.05$). The mean task completion time for **Ti+GT**, offering additional touch-based panning, increased compared to **Ti+G** (however, not significantly). The **Sc+G** and **To+G** achieved similar task completion times (no significant differences). Finally, the slightly slower task times in G1 for **Sc+G** compared to **Sc+M** could be due to the stabilized gaze cursor resulting in a dragging behind. Thus, more adaptive stabilizing approaches could speed up the results.

Quantitative user feedback. Results from the intermediate questionnaires are listed in Figure 1. In general, participants assessed all techniques, except for **Ti+GT**, very positive. Only for the *intuitiveness* the gaze-supported techniques were rated significantly lower than the familiar mouse condition **Sc+M**. **Ti+GT** received the lowest ratings in each category. However, **Sc+G** followed closely and was even rated better than **Sc+M** with respect to the *speed* in which actions could be issued and how well participants could orientate themselves after performing a movement (*spatial awareness*). Finally, **To+G** and **Ti+G** received similar ratings, except that **To+G** felt more intuitive than **Ti+G** (but less intuitive than **Sc+M** and **Sc+G**). After all pan-and-zoom techniques had been tested, participants were asked to rate how they liked each condition in the final questionnaire. Interestingly, **To+G** was preferred, closely followed by **Sc+G** and **Sc+M** (see Figure 1, *overall satisfaction*). Participants indicated that the gaze-supported panning is rather reasonable for small panning steps (M=4.30, SD=0.64) than for large ones (M=3.10, SD=0.94).

Qualitative user feedback. In general, the *gaze panning* was very positively assessed, especially in combination with the *gaze-pivot zooming*. Several participants particularly pointed out that they liked that the rough gaze data are sufficient for this type of interaction. The seamless combination of gaze-supported panning and zooming was praised. It was also positively remarked that there was

actually no need to look at the mobile device for issuing commands, as the designed touch and tilt techniques worked without the need to look away from the distant display. When getting the task to pan and zoom with **Sc+M** after having already tested a gaze panning condition, one participant actually mentioned that he would like to have the gaze panning, as it was more fun.

Participants explained that they did not like the *touch-based panning* in **Ti+GT**, because it was too complicated to coordinate tilting, gaze panning, and touch panning at the same time. Furthermore, it was distracting that as soon as touching the mobile screen in this condition, the gaze panning would be active. Instead, it was proposed to use a brief timeout before activating it, so that gaze and touch panning would not affect each other.

While all participants mentioned that **Sc+M** felt familiar and fast, some participants also indicated that the *scroll wheel zooming* sometimes felt cumbersome and that it would "take too long" although the maximum zoom speed was the same for all five conditions. The *touch-based zooming* was very positively assessed, as participants liked to be able to gradually adapt the zooming speed. However, a problem interrupting a smooth interaction, in particular for **To+G**, was an accidental leaving of the iPod's touch screen towards the inactive border region, because no haptic difference can be perceived. Thus, in order not to need to look away from the distant display, a better feedback should be given when the finger reaches the boundary of the touch screen. In this context, also additional feedback was desired for indicating the current zoom level. Finally, the *tilt-based zooming* was in general described as complicated and tiresome, as it required more physical effort compared to the other zooming techniques.

6 Discussion

All in all, the combination of gaze input and a mobile device for remote pan and zoom control was perceived as very promising. Users especially praised the gaze-directed pivot zoom and also found the gaze-directed panning easy to use. In this respect, the combination of a gaze-directed pivot zoom with a handheld integrating a touch gesture as for **To+G** or a scroll wheel as for **Sc+G** are particularly promising for a natural gaze-supported pan-and-zoom interaction. This is especially interesting when considering user contexts in which a mouse may not be available (e.g., sitting relaxed on your couch or for a flexible interaction with wall-sized displays). Instead of a scroll wheel, other physical input modalities could be combined with a handheld, such as a small switch lever allowing for small finger movements, haptic feedback, and the possibility to gradually adapt the moving speed depending on how much the lever is pushed. Considering that participants were already well trained with the mouse, further training with the gaze-supported techniques and better a gaze stabilization will increase performance.

In line with the work from Nancel et al. [2011], the scroll wheel and touch-based zoom achieved best results. However, further considerations for improving their combination with gaze input are required. This includes a well-thought-out design of the handheld to enable an interaction without the need to look away from the distant display, including improved haptic and auditory feedback. In addition, further improvements include speeding up large panning steps for gaze-supported input, for example, by incorporating flick gestures on the mobile device or quick gaze gestures. Although **Ti+GT** also aimed to achieve this by allowing both gaze and touch panning, the combination of two different panning modalities showed to confuse users. This indicates that rather simple modalities should be used or a more careful multimodal design for accessing each mode individually is required. Thus, to make it easier for users to clearly distinguish between the modes, distinct touch zones on the handheld or a physical mode switch could be used.

7 Conclusion

In this paper, we wanted to find out how users would enjoy and assess gaze-supported pan-and-zoom techniques using the example of a geographical information system. For this purpose, we described four combinations of gaze-directed panning with three different zooming modalities: (1) a mouse scroll wheel, (2) tilting a handheld device, and (3) touch gestures on a modern smartphone. These techniques and a control condition (using the mouse) were tested in a user study with ten participants. While the mouse-only condition yielded in the fastest task times, the combination of gaze-directed panning with a scroll wheel and with touch-based zooming was assessed very positively by the participants. Especially the possibility to zoom in towards the current point-of-regard was positively emphasized. These results are encouraging for further advancing gaze-supported techniques, particularly for user contexts in which traditional mouse input may not be available.

Acknowledgements

This research is supported by the German National Merit Foundation and the German Ministry of Education and Science (BMBF) project ViERforES-II (01IM10002B).

References

ADAMS, N., WITKOWSKI, M., AND SPENCE, R. 2008. The inspection of very large images by eye-gaze control. In *Proc. AVI'08*, ACM, 111–118.

BATES, R., AND ISTANCE, H. 2002. Zooming interfaces!: enhancing the performance of eye controlled pointing devices. In *Proc. Assets'02*, ACM, 119–126.

CASTELLINA, E., AND CORNO, F. 2008. Multimodal gaze interaction in 3D virtual environments. In *Proc. COGAIN '08*, 33–37.

FONO, D., AND VERTEGAAL, R. 2005. EyeWindows: evaluation of eye-controlled zooming windows for focus selection. In *Proc. CHI'05*, 151–160.

HANSEN, D. W., SKOVSGAARD, H. H. T., HANSEN, J. P., AND MØLLENBACH, E. 2008. Noise tolerant selection by gaze-controlled pan and zoom in 3D. In *Proc. ETRA'08*, ACM, 205–212.

LANKFORD, C. 2000. Effective eye-gaze input into windows. In *Proc. ETRA'00*, ACM, 23–27.

NANCEL, M., WAGNER, J., PIETRIGA, E., CHAPUIS, O., AND MACKAY, W. 2011. Mid-air pan-and-zoom on wall-sized displays. In *Proc. CHI'11*, ACM, 177–186.

STELLMACH, S., AND DACHSELT, R. 2012. Look & touch: Gaze-supported target acquisition. In *Proc. CHI'12*, ACM.

STELLMACH, S., STOBER, S., NÜRNBERGER, A., AND DACHSELT, R. 2011. Designing gaze-supported multimodal interactions for the exploration of large image collections. In *Proc. NGCA'11*, ACM, 1–8.

ZHANG, X., REN, X., AND ZHA, H. 2008. Improving eye cursor's stability for eye pointing tasks. In *Proc. of CHI'08*, ACM, 525–534.

ZHU, D., GEDEON, T., AND TAYLOR, K. 2011. Moving to the centre: A gaze-driven remote camera control for teleoperation. *Interact. Comput. 23* (January), 85–95.