GAZE AND FEET AS ADDITIONAL INPUT MODALITIES FOR INTERACTING WITH GEOSPATIAL INTERFACES

A. Çöltekin[a*], J. Hempel[b], A. Brychtova[a], I. Giannopoulos[c], S. Stellmach[d], R. Dachselt[e]

[a] Department of Geography, University of Zurich, Switzerland – arzu.coltekin@geo.uzh.ch, alzbeta.brychtova@geo.uzh.ch
[b] HMI specification at IAV GmbH– mail@julia-hempel.de
[c] Institute of Cartography and Geoinformation, ETH Zurich, Switzerland – igiannopoulos@ethz.ch
[d] Microsoft, Seattle, USA – stellmach@acm.org
[e] Interactive Media Lab, Technische Universität Dresden, Germany – dachselt@acm.org

Commission II, WG II/6

KEY WORDS: Interfaces, User Interfaces, Multimodal Input, Foot Interaction, Gaze Interaction, GIS, Usability

ABSTRACT:

Geographic Information Systems (GIS) are complex software environments and we often work with multiple tasks and multiple displays when we work with GIS. However, user input is still limited to mouse and keyboard in most workplace settings. In this project, we demonstrate how the use of gaze and feet as additional input modalities can overcome time-consuming and annoying mode switches between frequently performed tasks. In an iterative design process, we developed gaze- and foot-based methods for zooming and panning of map visualizations. We first collected appropriate gestures in a preliminary user study with a small group of experts, and designed two interaction concepts based on their input. After the implementation, we evaluated the two concepts comparatively in another user study to identify strengths and shortcomings in both. We found that continuous foot input combined with implicit gaze input is promising for supportive tasks.

1. INTRODUCTION

It has now been twenty years since Howard & MacEachren (1996) suggested that the design of interface tools would become a fundamental endeavor for working with geographic information (Çöltekin et al., 2009), and today, when working with Geographic Information Systems (GIS), we constantly interact with maps and an interface that surround these maps. To enable interactive map usage, GIS software packages offer many functions, e.g., to create geometric objects, edit labels or change the zoom level of the map. Only few of these functions can be 'mapped' (programmed, linked) directly to the limited number of mouse interactions, i.e., we cannot control *all* functions using a mouse. For example, click and drag functions can be mapped either to *drawing a line* or to *panning*. To solve those conflicts, switchable *modes* are introduced (e.g., you can choose the drawing or a panning mode). Current GIS software commonly use toolbars to switch between modes. However, mode switching interrupts the workflow and causes errors if we accidentally select the wrong mode. The indirect handling with toolbars can be time-consuming, annoying and can increase cognitive load. To minimize the mode switching through the toolbars, most GIS software packages provide shortcuts based on mouse and key combinations to perform frequently used tasks. However, it appears that many users do not explore shortcuts or face problems e.g., to undo the underlying functions because shortcuts do not have a visual representation or feedback (Cockburn et al., 2008). Therefore, we believe GIS software environments could benefit from additional interaction modalities. Furthermore, most GIS software feature *multiple linked views* or *multiple coordinated views*, through which we can study the problem at hand from multiple perspectives

(Bernasocchi et al., 2012). These multiple linked view displays could also benefit from additional ways to switch between the views other than the traditional mouse and keyboard interactions. To explore the potential benefits of additional input modalities in GIS-like software environments, in this paper, we introduce gaze and feet as *input channels* or *modalities* in addition to the traditional mouse and keyboard interactions. The resulting physical setup can be seen in Fig. 1.

Figure 1. Envisioned setup for gaze- and foot-input in a desktop context.

These additional input channels give us the opportunity to map selected tasks directly to gestural user input. Because of this direct mapping, gaze and feet interaction might be more intuitive than switching between modes using toolbar menus. Furthermore, user performance might increase because the

---

* Corresponding author

workflow will not be interrupted. In this project, we take a user-centered approach and first identify gestures based on expert input, and evaluate these novel interaction techniques in a user study after the implementation. As Fig 1 demonstrates, our implementation enables a seamless integration into a standard office environment.

As a proof of concept, we implemented the standard interaction tasks *zooming* and *panning*. These interactions are common in many domains and applications; and because our users are familiar with them, we can avoid individual differences in learning unfamiliar operations and focus on evaluating the input modalities themselves. Furthermore, because these operations exist in many domains, our results will be relevant to other graphic software beyond GIS environments. We believe our study is timely as eye tracking is becoming more accessible than ever before, both financially (Johansen, et al., 2015; Li et al., 2006) and computationally (e.g. Voßkühler, 2013).

2. RELATED WORK

Both gaze and foot input received significant interest in Human-computer Interaction (HCI) domain. This is motivated partly to give additional or alternative modalities for complex interfaces, and partly because input through gaze or foot would have meaningful applications for certain types of disabilities. However, combining gaze and foot input appears to have been rare (Göbel et al., 2013; Klamka et al., 2015). Klamka et al. (2015) also worked with zoom and pan. They demonstrated that their implementation outperformed the traditional mouse-only interaction and their users dominantly preferred the new combination of modalities including gaze and feet input. Their preliminary findings encourage further implementation and testing of these multimodal approaches. Our study complements their work by collecting gestures from GIS experts, thus specializes the input to the target user group. Furthermore, we implement the feet interaction through an off-the-shelf gaming board as opposed to Klamka et al. (2015)'s specialized pedal hardware. We believe using an off-the-shelf device makes our approach more accessible to anyone who wishes to reproduce our results, or experiment with the approach. While we did not encounter other examples of combined use of gaze *and* feet, below we give a brief overview of common GIS user interfaces as well as findings on navigation methods using gaze *or* feet.

2.1 Interacting with Geographic Information Systems

Because GIS offer a wide range of operations that involve creating, storing, manipulating, statistically analyzing and visualizing spatio-temporal data, user interfaces to geospatial applications are highly complex (Haklay, 2010; Schnürer, Sieber, & Çöltekin, 2015). Despite this complexity, currently, the provided communication channels in GIS are basic and mostly limited to keyboard and mouse input. The rest of the operations rely on interaction dialogues (Kuhn, 1999). The user interface of a GIS and the available interaction dialogues determine, to a great extent, how usable and useful such a system is (Egenhofer & Kuhn, 1999). Interaction dialogues evolved over the years, starting from command line interaction to the use of visual feedback, minimizing the necessity to memorize commands (Egenhofer & Kuhn, 1999). Today, GIS are equipped with "windows, icons, menu, pointer" (WIMP) style interfaces. WIMP interfaces were later expanded by additional modalities or even replaced by what has been termed "post-WIMP" interfaces. An often cited advantage of the post-

WIMP interfaces is that they allow users to focus on their task and not on the underlying technology (van Dam, 1997). Similarly, a common paradigm in HCI is to exploit what is considered 'natural' gestures. For example, utilizing emerging technologies, such as eye tracking, multimodal interactions can be enabled (Giannopoulos et al., 2012), and that multimodal interaction channels increase engagement (Egenhofer & Kuhn, 1999) as well as efficiency of users for map based tasks (Cohen et al., 2000). Increasing efficiency is an important goal in GIS environments, because often, large amounts of spatial, temporal and attribute information is processed and visualized in them (Li et al., 2015). Dealing with large amounts of information through complex interaction steps contributes to cognitive load for GIS users (Bunch & Lloyd, 2006). Thus, cognitive aspects of HCI are central and critically relevant for the use of GIS (Nyerges et al., 1995). One approach to reducing the cognitive load of the user is paying close attention to the visualization and interface design (e.g., Brychtová & Çöltekin, 2016; Çöltekin, et al., 2009), and another is cognitive engineering, by employing intuitive multimodal interaction channels developed based on human cognitive abilities, thus, minimizing the gap between the system and the user (Raubal, 2009). The need for effective, efficient, and intuitive interaction between the user and GIS becomes even more evident when looking at domains such as crisis management (Cai, et al., 2006) where functionality and the efficient interplay between the GIS and the user is crucial. In the crisis management domain, an early study by Fuhrmann et al. (2005) demonstrated that enabling multimodal interaction with GIS based on gesture and speech input leads even non-expert users to successfully achieve their goals. Furthermore, various researchers experimented with novel technologies, e.g., linking multi-touch interaction with GIS might have potential for collaborative work (Zadow et al., 2012). Similarly, the availability of depth sensors such as the Microsoft Kinect™ camera yield novel 'natural' interaction concepts for controlling virtual globes (Kamel Boulos et al., 2011).

2.2 Gaze Input

For many everyday tasks, humans have a coordinated action between the hands and their gaze, and in some situations we observe parallels between gaze behavior and mouse use (Çöltekin et al., 2014). Considering this, gaze appears to be an interesting candidate as an input modality and can potentially be used as a fast natural pointer. However, a (de)activation option is necessary to prevent unintended triggering of commands, i.e., the "Midas Touch" (Jacob, 1990). Gaze-contingent displays (GCDs) have been around for a long time (Duchowski & Çöltekin, 2007), and remain relevant in geospatial tasks (Bektas et al., 2015). Already more than 15 years ago, Zhai et al. (1999) experimented with using gaze input to support selection tasks. In recent years, Hansen et al. (2008) used eye tracking for zooming and panning, as well as for selecting objects in large information spaces. Stellmach et al. (2011) used a similar zooming technique, and their participants found it intuitive. Panning via gaze was implemented using 'active borders' concept by Hansen et al., (2008) and Stellmach et al., (2011); in which, if the user's gaze focuses on the border, the viewing frame is moved towards the corresponding direction. Stellmach et al. (2011) found that participants in general liked this interaction for short distances but not for covering large distances. Moreover, the study revealed that the borders need to be fairly large, and feedback for the active regions is desirable. Later, instead of using active borders, Stellmach & Dachselt (2012b) proposed a multimodal approach to distinguish between zooming and panning:

Touching or tilting a mobile device performed zooming, while panning was activated by a touch gesture on a mobile device and directed towards the user's gaze point. The evaluation in Stellmach & Dachselt (2012b)'s study suggests that participants found the gaze-directed panning easy to use, however, quick panning motions should be prevented to reduce disorientation and motion sickness. Around the same time, Netek (2011) suggested the concept of a gaze-controlled map viewer, in which certain dwell times would cause events on the map (triggering panning and zooming).

2.3 Foot Input

Even through foot movements are in general slower than hand movements (Hoffmann, 1991), they appear to be well-suited for tasks which do not require high accuracy and speed (Pakkanen & Raisamo, 2004). Because feet are used for locomotion in everyday life, foot input was investigated to develop appropriate navigation methods for virtual environments e.g., using treadmills (Darken et al., 1997), or a dance pad (Beckhaus et al., 2005). Besides the proposed dedicated devices such as the 'mole' instead of mouse (Pearson & Weiser 1986), and 'footmouse' (Simpson, 2013), a popular low cost device to capture foot input is the Wii Balance Board^{TM} by Nintendo, which recognizes the user's weight shift using four sensors. Haan et al. (2008) used the balance board to control 3D navigation in virtual environments and stated that it enables a smooth navigation. Moreover, Haan et al. (2008) noted that the balance board is capable of capturing continuous as well as discrete input, and can be used standing or seated. Schöning, et al. (2009) used the balance board combined with multi-touch input for interacting with spatial data in front of a large display: While multi-touch gestures were deployed for precise input, continuous foot input was used to cover large distances, e.g., zooming was possible by pointing with the hand on a location of the map and leaning forward. The initial evaluation in Schöning, et al. (2009)'s study suggests that the technique enables a smooth navigation; however, the balance board limits the freedom of movement in front of the large display.

3. CONCEPT DEVELOPMENT

The literature review above suggests that gaze as well as foot input provide potentially useful interaction opportunities to enable navigation in geospatial visualizations. As it is with the hand-eye coordination, our eyes and feet are coordinated to some degree (Mikaelian, 1970, Hollands & Marple-Horvat 2001). Thus combining gaze, hand and feet input lends itself to an interesting experiment -- the coordination between the three can help or hinder the user. However, based on earlier studies, a multimodal approach seems to be rather complementary than competitive in terms of cognitive resources, promising to overcome the weaknesses of one another.

3.1 Interaction Concept

To identify suitable gestures for zooming and panning with geovisualizations, we conducted a preliminary user study with six expert GIS users. The participants were asked to imagine gestures combining gaze and feet input to zoom and pan. To support the participants' imagination of the future applications, we chose a setup similar to the prospective working environment: the workplace contained a desktop computer which runs the application ArcMap as a typical example of GIS software. An external eye tracker was situated under the

monitor and a Wii Balance Board under the table (neither was switched on at this point). The participants were asked to sit with their feet on the balance board. Foot gestures and comments were recorded with a camera. We asked the participants to imagine a way to perform zooming and panning using feet and gaze input. In general, they anticipated that they would prefer foot input to gaze, because they were afraid that gaze input might disturb the normal workflow. For panning, most participants preferred continuous foot input. They proposed to press the feet to one direction of the balance board in order to move the viewing frame in the corresponding direction. Thus, the direction of the user's weight shift is mapped directly to the direction of the panning motion. For zooming, the results of the preliminary study suggest pressing both balls of the feet/heels down. The resulting change of the scale level should be continuous and directed towards the user's gaze point. Because the user-elicited mappings (i.e., mapping the gesture to input device) indicate a conflict between zooming in/out and panning up/down, we designed two different approaches: *Surfing*, and *Walking*.

3.2 Surfing Concept

The Surfing Concept is based on the user-elicited 'mappings'. Zooming in and zooming out were mapped to shifting the weight forward/backward while panning was mapped to shifting the weight to the desired direction on the balance board. To solve the conflict between zooming in/out and panning up/down, the visualization is separated in a central zoom area surrounded by panning areas as proposed by Hansen et al. (2008), and illustrated in Fig. 2. Panning is active if the user's gaze is fixated on a panning area while zooming is active if the user focuses the central area. The velocity of zooming and panning depends on the amount of conducted pressure.
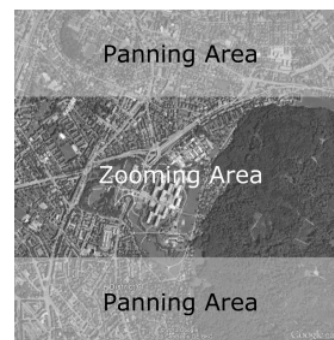

Figure 2. Surfing concept: Geographic visualization with panning areas.

The more users press their feet down, the faster the resulting movement becomes. Examples of user commands are shown in Fig. 3 and 4. As the preliminary study suggested, zooming was directed towards the user's gaze point, i.e., we used a "fixed point zoom" (Furnas & Bederson, 1995). This means user's point of regard remained at a fixed location on the display while zooming. In contrast to central zoom, this zoom trajectory should allow easy tracking of the target while zooming.

3.3 Walking Concept

The Walking Concept solves the conflict between zooming and panning up/down by using a discrete taps for zooming. Thus, tapping with the ball of one foot performs zooming in (Fig. 5). Analogously, tapping with the heel (i.e., at the area 'behind' on

the board, instead of 'front' – one *can* use another part of the foot if/when appropriate, what matters is the location of the tap) of one foot causes a zoom out. Each tap causes a change of the zoom level with a discrete step. Again, the area, which is focused by the user remains at a fixed location in the display while zooming. As in the Surfing Concept, panning is mapped to shifting the weight on the balance board (Fig. 6 and 7).

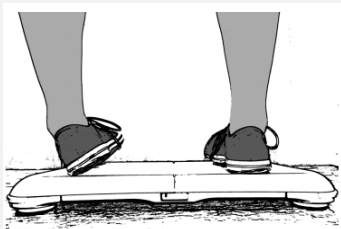
Figure 3. Surfing concept: Pan up or zoom in.

Figure 4. Surfing concept: Pan left.

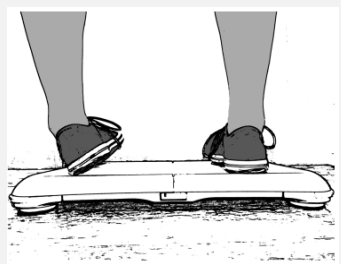Figure 5. Walking concept: Zoom in.

Figure 6. Walking concept: Pan up.

Figure 7. Walking concept: Pan left.

The velocity of panning is not constant but rate-based. It is mapped to the amount of pressure expended by the user. That means, the harder a user presses her or his feet, the faster becomes the motion of the map.

4. IMPLEMENTATION AND EVALUATION

4.1 Implementation

Our implementation was based on an existing eye tracking framework, written in VB.Net and uses the Tobii SDK 2.0 to handle gaze input from a Tobii Eye Tracker (Giannopoulos et al., 2014). We adapted this framework to our purposes by adding a geographic visualization using the Google Earth Plug-in and a connection to the Wii Balance Board (connected to the framework using Bluetooth). The calibration was done based on the initial values of the pressure sensors (top-left, top-right, bottom-left, bottom-right) which depend on the weight and seating position of the user. After starting the application, we polled the values of the pressure sensors every 50ms. Based on the calibration, the values were normalized. If the resulting pressure values exceed a certain threshold, an action is triggered. For gaze recognition, we used the algorithm by with 30mm radius around a fixation point and a threshold of 400ms. To provide feedback about the current gaze position, the gaze point was visualized using the cursor position.

4.2 User Study (Evaluation)

We conducted a controlled user study to evaluate the two concepts we developed and implemented, and to receive feedback to further improve the prototype.

4.2.1. Experimental Design: The study was designed as a within-subject factorial experiment. The main independent variables were our two concepts as described above (*Surfing* and *Walking*), and as control variables, we used two common input modalities (*Toolbar* and *Mouse*). All participants completed all tasks using both the Walking and Surfing concepts in a systematically rotated order; as well as the traditional mouse input using buttons and wheel, and toolbar icons *Zoom in/out* and *Pan*. Thereby, we distinguished between mouse interaction common in Google Earth (no use of visual icons) and mouse interaction using a toolbar where a visual icon was used. Consequently, four concepts were tested. We measured *time on task,* obtained *overall satisfaction* scores, and asked participants to provide *qualitative feedback*.

4.2.3. Tasks: Experiment included basic interaction tasks (zoom, pan, go to full-extent) and combined interaction tasks (e.g., pan *and* zoom). The task setting was to "fit a red circle into a grey border" (Fig. 8). Depending on the location of the grey border, the participants needed to zoom or pan in a given direction or combine both interaction commands.

4.2.2. Participants: The study was conducted with 16 voluntary participants (9 females, 7 males), aged 20-40 (and one over 40). Most of them were students or employees at the Department of Geography at the University of Zurich. Consequently, they rated their experience using digital maps very high (mean = 5.8 on a 7-point Likert-scale). The experience using GIS software was rated slightly lower compared to digital maps but was also high (mean = 5). In contrast, the participants' experience using eye-tracking applications and the Wii Balance Board was low in general.

Participants rarely or never used eye-tracking applications before (mean = 2.9), except one. Their experience with the Wii Balance Board was even less than in eye-tracking (mean = 1.9). 10 out of 16 participants never used the Wii Balance Board before, and only one stated that he/she was an experienced user.
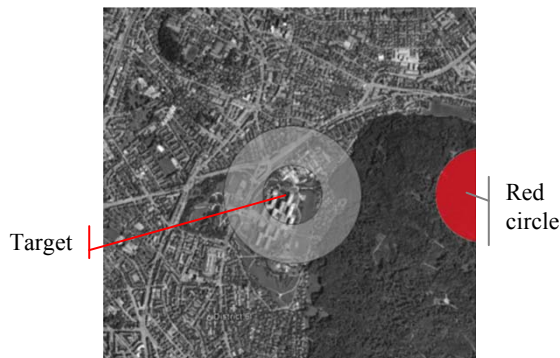

Figure 8. Task setting for panning: drag the red circle inside the grey border (fixed position).

4.2.4. Procedure: After welcoming the participants, we asked them to fill in an 'entry questionnaire'. This questionnaire contained the demographic questions (age, gender, education, etc.) and statements concerning the participants' experiences in working with digital maps (Google Maps) and GIS software (ArcMap) as well as their experiences in using the Wii Balance Board and eye-tracking applications. We also asked them to provide us how frequently they use traditional interaction methods (toolbar, mouse buttons and keyboard shortcuts) while working with Google Maps and ArcMap. The frequency of using interaction methods was rated using a 7-Point Likert-scale (1=never, 7=very often). After filling in the entry questionnaire, the interaction methods were tested one after the other in a systematically rotated order. In each case, the participants first received the description of the interaction method and had time to try it out in order to get familiar with the interactions. After the familiarization, the study started. At the end a final questionnaire was handed out in order to assess the participants' satisfaction with all tested methods. We asked participants to provide rating concerning their overall satisfaction using a 7-Point Likert-scale (1=strongly unsatisfied, 7=strongly satisfied). Each trial lasted about one hour.

5. RESULTS

5.1 Use of Traditional Interaction Techniques

The difference in the use of interaction techniques in GIS and Google Maps are shown in Fig. 9.
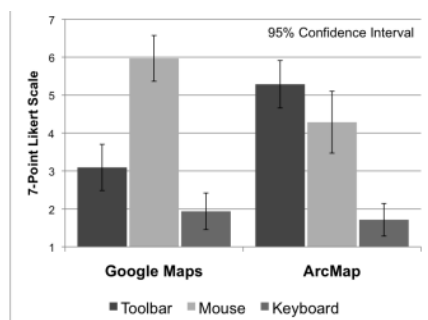

Figure 9. Frequency of use of traditional techniques to perform map navigation tasks (1=never using, 7= very often using).

Most participants stated that they use standard mouse interactions, while they use the toolbar rarely. In contrast, most participants stated that they perform zooming and panning in ArcMap predominantly with the toolbar. Especially for panning, wheel- and button-based mouse interactions are used rarely (mean=3.4) while the toolbar with visual icons is used a lot (mean= 5.5). Keyboard shortcuts are in general used rarely for interaction (Fig. 9).

5.2 Response times

As Fig 10 shows, participants completed the interaction tasks faster using the mouse and toolbar than using feet and gaze (surfing and walking concepts). This is clearly not surprising given how familiar they were with the mouse use.
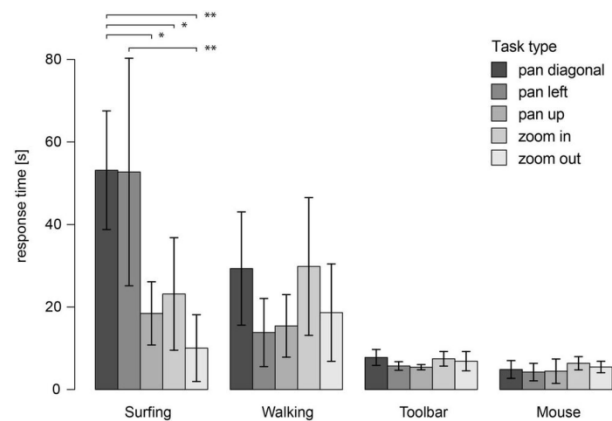

Figure 10. Mean response times. * p< .05, ** p < .01. Error bars indicate 95% confidence interval.

Median times, regardless of the task type, were; 18.39s for Surfing, 15.03s for Walking, 5.43s for Mouse and 9.16s for the Toolbar. A Kruskal-Wallis test revealed dependency between the response time and the input method (H=204.78, df=3, p<.001). Dunn's post-hoc pair-wise comparison showed that the Surfing and Walking did not differ (p=.39), while all other differences remained significant (all p=<.001). Fig. 10 shows that the mean task completion times did not differ between Surfing, Walking and Toolbar for the tasks panning left, zoom in and zoom out. However, for panning up and panning diagonal, a difference was detected in favor of Walking rather than Surfing. We further studied the dependency between the input modalities and the task types (zoom in, zoom out, pan diagonal, pan left, pan right): the Walking concept and the Toolbar input were not dependent on the task type (H=8.31, df=4, p=0.08, respectively H=4.41, df=4, p=.35), but Surfing and Mouse were dependent on the task type (Surfing H=36.36, df=4, p<.001; Mouse H=13.56, df=4, p=.008). While Surfing, participants had the longest response time for panning diagonal, and to the left, compared to other tasks (and also to other input concepts). Surprisingly, with Surfing the response time did not depend on the modality (H=5.68, df=3, p=.13), suggesting that this task was as easy with the Surfing concept as it was with the traditional input modalities (Mouse and Toolbar).

5.3 Subjective Evaluation

The results of the final questionnaire concerning the overall satisfaction are presented in Fig. 11. As can be seen in Fig. 11, participants rated the mouse interaction as most satisfying overall, followed by the Toolbar. However, differences between the Toolbar, Surfing and Walking are not too large.
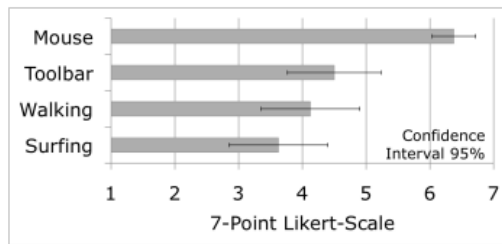
Figure 11. Overall satisfaction (1=strongly unsatisfied, 7=strongly satisfied).

The participants gave a lot of qualitative feedback using the open-ended questions and through free verbal comments during the study. Comparing Walking and Surfing concept, they felt that the Walking Concept were more suitable for zooming in and out, while the Surfing Concept was more satisfactory for panning commands. Using the Surfing Concept, participants especially praised the continuous zooming. They stated that they found it fast and easy to zoom in and out of the map. Few participants reported that the speed, especially to zoom in, was too fast. Moreover, zooming towards the gaze point was assessed positively. Panning by shifting the weight was reported as "easy and quick". One participant even noted that she/he found the panning faster than using a mouse. The main point of criticism was the distinction between zooming and panning up/down using gaze input.

Many participants faced problems to focus on the "active borders" while panning, because it conflicts with visually tracking the target. Participants expressed that precise panning up/down was very difficult without the possibility to look at the target. These difficulties are reflected by the measured response times (Fig. 10). Participants were more than 30s slower when panning up with Surfing concept, than Walking. Thus, this concept may not be ideal for geospatial visualizations. Diagonal panning was the most problematic task for both foot-eye navigation techniques; however while Surfing, participants were more than 20s slower than while Walking. For panning left, no differences between foot-eye concepts were observed. On the other hand, many participants focused on the target while panning towards the centre of the display, and as a result, zooming and panning operations were conflicted, causing confusion for the users.

In general, the Walking Concept was reported as "more intuitive" than Surfing. Participants especially praised the panning as "easy and intuitive". Two participants noted that they liked the panning because it worked the same way in all directions. The distinction between panning and zooming was reported to be easier than using the Surfing method. However, according to our participants, zooming by tapping was more physically demanding than pressing both feet down. Moreover, many participants faced problems caused as they confused the functions mapped to tapping. Only few participants liked the direction of zooming towards the gaze point, while majority of the participants did not seem to fully comprehend (or learn) that the zoom was directed towards the gaze point, suggesting that this operation has a high learning curve.

6. DISCUSSION

In the course of this work, we developed two novel interaction methods using gaze and feet input to navigate in geographic visualizations based on cognitive assumptions of a certain level

of coupling between eyes and feet (Hollands & Marple-Horvat, 2001; Hollands et al., 2002; Mikaelian, 1970), and previous work that motivates such solutions from a practical perspective, i.e., to unburden the hands, or to aid them. We first conducted a requirements study and designed our concepts accordingly, after which we implemented the alternative modalities and conducted a user study with expert GIS users to evaluate the proposed concepts. Our findings show that the users did best with the more traditional modalities, especially the mouse. This is not surprising, given that the utilized input devices (the eye-tracker and the Wii Balance Board) were rarely used by most participants. These results may also be explained by the fact that we needed to calibrate the two new input devices, and this caused problems, especially for very small or tall participants due to limitations of the physical setup. Because of these, results should be interpreted carefully to distinguish between conceptual problems and technical ones. Overall, the results of this initial study were somewhat mixed, but we have reasons to believe that the new methods are promising.

Regarding the two novel interaction methods we proposed (Surfing and Walking), the user study revealed advantages and shortcomings in both. The Surfing method overall facilitated a faster zooming than Walking, while the participants performed better with the Walking method in panning. The qualitative user feedback confirmed these findings. While the continuous zooming (pressing both feet) was said to be fast and easy, the discrete zooming (tapping one foot) was said to be less smooth and physically more demanding than the continuous input. For panning, both methods used continuous feet input. However, the Walking method was faster than the Surfing method and was assessed more positively by the participants. The qualitative feedback pointed out that this difference might be explained based on the distinction between zooming and panning. While panning up/down using the Surfing concept needed to be activated by looking at the active borders, no activation was necessary using the Walking method. Thus, with the Walking method, the feet interaction was symmetric for all directions, and participants noted this as a positive feature, as the mapping from feet input to panning direction was clear to all participants without the need to look at the description again. In general, participants found panning 'easy and intuitive' with the Walking method. These results and qualitative feedback confirm the previous studies (e.g., Klamka et al., 2015, Stellmach & Dachselt 2012b), thus we confirm that the tested interaction modalities lead to promising results also with GIS experts, and the specific hardware configuration we used.

The use of gaze as an input modality was assessed mixed, too. When the gaze input was used for determining the direction of zooming, participants found it helpful, which might be an expression of the inherent 'foot-eye coordination' (Hollands & Marple-Horvat, 2001). At this point, it is worth noting that gaze-directed zooming was more effective using continuous zooming than discrete zooming. In contrast to determining the gaze direction, the use of gaze input to distinguish between zooming and panning was assessed mostly negatively. Because eyes are not always consciously controlled, accidentally mixing up between functions is quite possible, thus interaction must be designed carefully. Furthermore, participants reported that the explicit gaze input required a lot of concentration to control the eyes and hampered an unrestricted viewing of the visualization. As a result, gaze input should be limited, e.g., to implicit use, in order to release users from the burden to control their eyes consciously. Concluding from these results, continuous feet input is better suited for map navigation than discrete feet input.

However, the mapping needs to be direct so that no activation or mode switching is necessary. Under these assumptions, we think that continuous feet input combined with implicit gaze input is very promising. Even in this early state of the prototype, the task completion times were not always significantly lower than the mouse input in the user study. Thus, feet and gaze input might become equally fast as mouse input by improving calibration of speed and sensitivity, as well as user training.

7. CONCLUSIONS AND OUTLOOK

The results of our evaluation suggest that the combination of gaze and foot interaction are promising methods as complementary input modalities to traditional mouse and keyboard interactions, or in some cases even an alternative to commonly used toolbar icons. With careful design and proper testing, they could become helpful additions to near future offices.

As noted before, in this study most participants were familiar with mouse and keyboard interactions as well as toolbar icons as opposed to these new modalities we explored. This is likely to be the case also in the larger population. We believe the user satisfaction with gaze and foot as input modalities might increase with further familiarization (remember, e.g., when we used mouse the first time, it was not straightforward either[1]). Our findings in the user study suggest that especially foot input mapped to manipulation tasks like zooming and panning is recommendable. Using feet as a shared, alternative or complementary input modality may also have implications for office ergonomics, i.e., we might use our hands a little less, which could mean lesser risk for e.g., repetitive strain injury, and potentially help people with other disabilities. Of course, conversely, it may also create new ergonomic problems, thus should be tested specifically for long-term use as well.

Gaze input appears to be well suited to *implicit* pointing e.g., for gaze-directed zooming. On the other hand, based on our observations in this study, we believe the use of explicit gaze input should be avoided, or designed very carefully as it might have cognitive drawbacks, e.g., it might 'overcharge of the visual channel' (Broy & Rümelin, 2012). In other words, our visual system processes a lot of information in executing geospatial tasks, therefore it should be carefully considered what is a sensible gaze-interaction paradigm and validated in future experiments. To further improve the methods, other foot input devices, like interactive shoes or multi-touch floors might be worth considering for adding interaction options and raising the comfort. At this point we are convinced that if the users are able to switch on and off the feet and gaze as input modalities (it is important to turn them off to minimize the possible frustration with 'Midas touch'), they would adapt and benefit from having these as alternatives or complementary options to our current practices.

ACKNOWLEDGEMENTS

REFERENCES

Beckhaus, S., Beckhaus, S., Blom, K. J., & Haringer, M. (2005). Intuitive, hands-free travel interfaces for virtual environments. *IEEE VR Workshop: New Directions in 3D User Interfaces*, 57–60.

Bektas, K, Çöltekin, A., Krüger, J., Duchowski, A. T. (2015). A testbed combining visual perception models for geographic gaze contingent displays. *Eurographics Conference on Visualization (EuroVis) - Short Papers*.

Bernasocchi, M., Çöltekin, A., Gruber, S. (2012). An open source geovisual analytics toolbox for multivariate spatio-temporal data for environmental change modeling. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *1-2*(2), 123–128.

Broy, N., & Rümelin, S. (2012). Natural visual user interfaces– beyond input modalities. *Adjunct Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '12)*, 61–63.

Brychtová, A., & Çöltekin, A. (2016). The effect of spatial distance on the discriminability of colors in maps. *Cartography and Geographic Information Science*, 1–17.

Bunch, R. L., & Lloyd, R. (2006). The cognitive load of geographic information. *The Professional Geographer*, *58*(2), 209–220. doi:10.1111/j.1467-9272.2006.00527.x

Cai, G., Sharma, R., MacEachren, A. M., & Brewer, I. (2006). Human-GIS interaction issues in crisis response. *International Journal of Risk Assessment and Management*, *6*(4), 388–407.

Cockburn, A., Karlson, A., & Bederson, B. B. (2008). A review of overview+detail, zooming, and focus+context interfaces. *ACM Computing Surveys*, *41*(1), 1–31.

Cohen, P., McGee, D., & Clow, J. (2000). The efficiency of multimodal interaction for a map-based task. In *Proceedings of the Sixth Conference on Applied Natural Language Processing* (pp. 331–338). Stroudsburg, PA, USA: Association for Computational Linguistics.

Çöltekin, A., Demsar, U., Brychtova, A., & Vandrol, J. (2014). Eye-hand coordination during visual search on geographic displays. In *Proceedings of the 2nd International Workshop on Eye Tracking for Spatial Research, GIScience2014*. Vienna, Austria.

Çöltekin, A., Heil, B., Garlandini, S., & Fabrikant, S. I. (2009). Evaluating the effectiveness of interactive map interface designs: A case study integrating usability metrics with eye-movement analysis. *Cartography and Geographic Information Science*, *36*(1), 5–17.

Darken, R. P., Cockayne, W. R., & Carmein, D. (1997). The omni-directional treadmill: a locomotion device for virtual worlds. In *Proc. UIST 1997* (pp. 213–221). ACM Press.

Duchowski, A. T., & Çöltekin, A. (2007). Foveated gaze- contingent displays for peripheral LOD management, 3D visualization, and stereo imaging. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *3*(4), 1–18.

Egenhofer, M., & Kuhn, W. (1999). Interacting with GIS. In P. Longley, M. Goodchild, D. Maguire, & D. Rhind (Eds.), *Geographical Information Systems: Principles, Techniques, Management and Applications* (2nd ed., pp. 401–412). New York, USA: Wiley.

---

[1] "Grandma vs. mouse" https://www.youtube.com/watch?v=Hwfgwn1yB8Q

Fuhrmann, S., MacEachren, A., Dou, J., Wang, K., & Cox, A. (2005). Gesture and speech-based maps to support use of GIS for crisis management: A user study. In *AutoCarto* (pp. 1–7).

Furnas, G. W., & Bederson, B. B. (1995). Space-scale diagrams: Understanding multiscale interfaces. In *Proc. CHI 1995* (pp. 234–241).

Giannopoulos, I., Kiefer, P., & Raubal, M. (2012). GeoGazemarks : providing gaze history for the orientation on small display maps. In *Proceedings of the 14th International Conference on Multimodal Interaction*.

Giannopoulos, I., Schöning, J., Krüger, A., & Raubal, M. (2014). Attention as an input modality for Post-WIMP interfaces using the viGaze eye tracking framework. *Multimedia Tools and Applications*, 1–17.

Göbel, F., Klamka, K., Siegel, A., Vogt, S., Stellmach, S., & Dachselt, R. (2013). Gaze-supported foot interaction in zoomable information spaces. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13* (p. 3059). New York, New York, USA: ACM Press.

Haan, de G., Griffith, E. J., & Post, F. H. (2008). Using the Wii Balance Board as a low-cost VR interaction device. In *Proc. VRST 2008* (pp. 289–290). ACM Press.

Haklay, M. (2010). Interacting with geospatial technologies. In *Interacting with Geospatial Technologies* (p. 128). John Wiley & Sons, Ltd.

Hansen, D. W., Skovsgaard, H. H. T., Hansen, J. P., & Møllenbach, E. (2008). Noise tolerant selection by gaze-controlled pan and zoom in 3D. In *Proc. ETRA 2008* (pp. 205–212).

Hoffmann, E. R. (1991). A comparison of hand and foot movement times. *Ergonomics*, *34*(4), 397.

Hollands, M. A., & Marple-Horvat, D. E. (2001). Coordination of eye and leg movements during visually guided stepping. *Journal of Motor Behavior*, *33*(2), 205–216.

Hollands, M. A., Patla, A. E., & Vickers, J. N. (2002). Look where you're going!: Gaze behaviour associated with maintaining and changing the direction of locomotion. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *143*(2), 221.

Howard, D. L., & MacEachren, A. M. (1996). Interface design for geographic visualization: Tools for representing reliability. *Cartography and Geographic Information Science*, *2*(23), 59–77.

Jacob, R. J. K. (1990). What you look at is what you get: eye movement-based interaction techniques. In *Proc. CHI 1990* (pp. 11–18). ACM Press.

Johansen, S. A., Agustin, J. S., Tall, M., & Skovsgaard, H. (2015). The Eyetribe - The world's first affordable eye tracker with full SDK.

Kamel Boulos, M., Blanchard, B., Walker, C., Montero, J., Tripathy, A., & Gutierrez-Osuna, R. (2011). Web GIS in practice X: a Microsoft Kinect natural user interface for Google Earth navigation. *International Journal of Health Geographics*, *10*(1), 1–14.

Klamka, K., Siegel, A., Vogt, S., Göbel, F., Stellmach, S., & Dachselt, R. (2015). Look & Pedal. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction - ICMI '15* (pp. 123–130). New York, New York, USA, New York, USA: ACM Press.

Kuhn, W. (1999). Paradigms of GIS use. In *5th International Symposium on Spatial Data Handling* (Vol. 1, pp. 91–103).

Li, D., Babcock, J., & Parkhurst, D. (2006). openEyes: a low-cost head-mounted eye-tracking solution. ... *of the 2006 Symposium on Eye Tracking ...*, *1*(March), 27–29.

Li, S., Dragicevic, S., Castro, F. A., Sester, M., Winter, S., Çöltekin, A., … Cheng, T. (2015). Geospatial big data handling theory and methods: A review and research challenges. *ISPRS Journal of Photogrammetry and Remote Sensing*.

Mikaelian, H. H. (1970). Adaptation to rearranged eye-foot coordination. *Perception & Psychophysics*, *8*(4), 222–224.

Netek, R. (2011). Possibilities of contactless control of web map applications by sight. In *Geoinformatics FCE CTU* (pp. 55–72).

Nyerges, T., Mark, D., Laurini, R., & Egenhofer, M. (1995). Cognitive aspects of HCI for GIS: An introduction. In T. Nyerges, D. Mark, R. Laurini, & M. Egenhofer (Eds.), *Cognitive Aspects of Human-Computer Interaction for Geographic Information Systems* (Vol. 83, pp. 1–8). Springer Netherlands.

Pakkanen, T., & Raisamo, R. (2004). Appropriateness of foot interaction for non-accurate spatial tasks. In *Ext. Abstracts CHI 2004* (pp. 1123–1126). ACM Press.

Raubal, M. (2009). Cognitive engineering for geographic information science. *Geography Compass*, *3*(3), 1087–1104.

Schnürer, R., Sieber, R., & Çöltekin, A. (2015). The next generation of atlas user interfaces: A user study with "digital natives." In *Modern Trends in Cartography - Lecture Notes in Geoinformation and Cartography* (pp. 23–36).

Schöning, J., Daiber, F., Krüger, A., & Rohs, M. (2009). Using hands and feet to navigate and manipulate spatial data. In *Ext. Abstracts CHI 2009* (pp. 4663–4668). ACM Press.

Simpson, R. C. (2013). *Computer Access for People with Disabilities: A Human Factors Approach*. CRC Press.

Stellmach, S., & Dachselt, R. (2012). Designing gaze-based user interfaces for steering in virtual environments. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 131–138).

Stellmach, S., & Dachselt, R. (2012b). Investigating gaze-supported multimodal pan and zoom. In *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA '12* (p. 357). New York, New York, USA: ACM Press.

Stellmach, S., Stober, S., Nürnberger, A., & Dachselt, R. (2011). Designing gaze-supported multimodal interactions for the exploration of large image collections. In *Proc. NGCA 2011*. ACM Press.

van Dam, A. (1997). Post-WIMP user interfaces. *Commun. ACM*, *40*(2), 63–67.

von Zadow, U., Daiber, F., Schöning, J., & Krüger, A. (2012). GeoLens: multi-user interaction with rich geographic information. *Proc. DEXIS 2011*, 16–19.

Voßkühler, A. (2013). OGAMA (OpenGazeAndMouseAnalyzer). Freie Universität Berlin.

Zhai, S., Morimoto, C., & Ihde, S. (1999). Manual and gaze input cascaded (MAGIC) pointing. In *Proc. CHI 1999* (pp. 246–253). ACM Press.