Gaze-supported Foot Interaction in Zoomable Information Spaces

Fabian Göbel
TU Dresden
Dresden, Germany
fabian.goebel@mailbox.tu-dresden.de

Stefan Vogt
TU Dresden
Dresden, Germany
stefan.vogt1@mailbox.tu-dresden.de

Konstantin Klamka
TU Dresden
Dresden, Germany
konstantin.klamka@mailbox.tu-dresden.de

Sophie Stellmach
Interactive Media Lab
TU Dresden
Dresden, Germany
stellmach@acm.org

Andreas Siegel
TU Dresden
Dresden, Germany
andreas.siegel@mailbox.tu-dresden.de

Raimund Dachselt
Interactive Media Lab
TU Dresden
Dresden, Germany
dachselt@acm.org

Abstract

A multimodal combination of gaze and foot input is highly promising for supporting manual interactions using, for example, mouse and keyboard. This is particularly interesting for simultaneously performing primary (e.g., object selection or manipulation) and secondary tasks (e.g., pan and zoom) in zoomable information spaces. While our eye gaze is ideal to quickly indicate a user's current point of interest, foot interaction is well suited for parallel hand-free input controls, for example, to quickly confirm an action. This allows for using gaze input in a subtle and unobtrusive way, while still maintaining a fast and convenient interaction. Motivated by this, we present several alternatives for multimodal gaze-supported foot interaction to pan and zoom in a desktop computer setup. With this, we contribute the novel approach of seamlessly combining gaze and foot input for a more convenient interaction.

Author Keywords

Multimodal interaction, foot, gaze, eye tracking, navigation, pan, zoom

ACM Classification Keywords

H.5.2 [Information interfaces and presentation]: User Interfaces: Input devices and strategies.

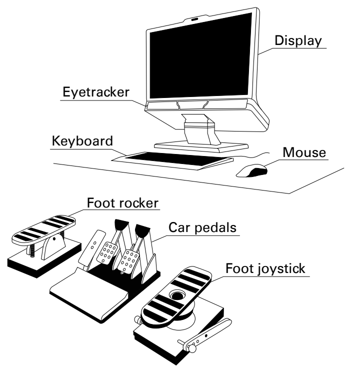

Figure 1: Extended desktop workplace with our proposed input devices.

Introduction and Motivation

In the WIMP ("Windows, Icons, Menus, Pointer") world, interaction possibilities are usually limited to mouse and keyboard input. This is also the case for zoomable information spaces that are used in a variety of application areas, such as image editing or geographical information systems. For such information spaces, complex interaction tasks often require to switch between different interaction modes or input devices which may interrupt the workflow. In this context, we can distinguish complex interaction tasks into primary tasks (e.g., manipulating a selected object) and secondary tasks (e.g., panning and zooming). We understand secondary tasks as actions that may support a currently performed primary action (cf. [4]). For example, while editing an image, the user may want to get a closer look at details in it by further zooming in.

To support performing primary and secondary tasks simultaneously and thus without interrupting the workflow, we propose a novel combination of post-WIMP input devices [9], namely gaze and foot input, with standard desktop workplaces. With this, we can extend the way how we commonly interact with our computers. For primary tasks, that often require high input precision, we consider manual input as most appropriate. Since secondary tasks may act as supporting activities for a currently performed primary interaction task, they should be executable with low mental and physical effort to avoid unnecessarily distracting a user from his/her primary task. In this context, gaze and foot input show high potential as supporting input channels, but are rarely used for human-computer interaction so far. First, gaze input is suitable as a fast, implicit and coarse pointing modality. Secondly, foot input is well suited to supplement implicit gaze input as it allows for parallel explicit input controls, for example, to confirm an action or to specify a zooming speed. Thus, this input combination allows for addressing prevalent challenges associated with gaze interaction, such as the *Midas Touch* problem which is considered as one of the major challenges for gaze-based interaction [3].

In this paper, we focus on the question how common pan and zoom actions can be conveniently performed through gaze-supported foot interaction in a desktop computer setup as illustrated in Figure 1. For this, we discuss several variants of foot-based input that allow for different ways how to map pan and zoom actions to gaze and foot controls. In the following, we first briefly discuss related work on foot- and gaze-based interaction. Then we describe our developed gaze-supported foot input and associated design considerations. Since this paper can only provide a broad overview of our ongoing activities, it will be closed with a discussion of future work.

Related Work

Our eye gaze is one of the fastest possible pointing methods, since our gaze reaches a target prior to a manual pointer without even thinking about it [7, 10]. Several multimodal gaze-supported pan and zoom approaches exist, e.g., with additional head movements [1] or a touch-sensitive handheld [7, 8]. Adams et al. [1] compare four different pan and zoom input techniques including gaze-based panning with zooming by clicking a certain mouse button, moving the head towards or away from the screen, and a gaze dwell-based activation. Stellmach and Dachselt [7, 8] investigate five pan and zoom variants in combination with a mouse scroll wheel, single touch gestures, and tilting of a handheld smartphone. In particular pivot zooming [8], the implicit use of gaze to indicate where to zoom in was highlighted by participants in their user study. However, quick panning motions via gaze input should be avoided as this may cause disorientation and motion sickness for the user [7].

Already in the 1980s, Pearson and Weiser [5] propose using foot input as supporting input to overcome interruptions of the user's workflow due to shifting the hand between mouse and keyboard. Pakkanen and Raisamo [4] investigate the appropriateness of foot input for non-accurate spatial tasks and propose to assign supporting interaction tasks, such as scrolling, moving or resizing objects to the feet. They conclude that *"feet are suitable for the secondary tasks that do not require high accuracy or execution times"* [4]. Several works address multimodal foot input, e.g., with multi-touch gestures (e.g., [2, 6]), but none with gaze input. While Pearson and Weiser [5] have already considered the promising potential for combining gaze and foot input, further investigations have been hindered by the extensive costs and inconvenience of eye tracking devices for human-computer interaction so far.

Gaze-supported Foot Interaction Concepts

We investigate combinations of gaze and foot input as shown in Figure 1 to seamlessly pan and zoom. For this, we have developed three prototypes that combine gaze input with different variants of foot input, such as foot pedals, a foot rocker, and a foot joystick. First, we describe some basic considerations for the design of our gaze-supported foot interaction techniques and discuss how we can address challenges for gaze-based interaction. Then we briefly introduce some of the design goals that we have pursued for our different variants of foot input.

For our design, we assume an ordinary desktop computer setup with a single user working at the desk which implies a seated position. As mentioned before, the basic idea is to free the hands for other manual input, such as mouse and keyboard controls which are commonly used in such contexts. Other, more advanced manual input modalities could be likewise considered for a combination with gaze-supported foot-interaction, such as freehand gestures, touch, or digital pen input. In this paper, we do not target a specific type of manual input and focus on a suitable combination of gaze and foot input instead. We use multiple pointers (or cursors respectively) for primary manual and for secondary interaction tasks. For example, this avoids a conflict between the mouse cursor required for manipulating content and a secondary pointer to indicate where to zoom in or in which direction to pan. Hence, the mouse cursor is *not* influenced by gaze or foot input.

Several challenges have to be taken into account for the design of gaze-based input controls. First, we need to address the *Midas Touch* problem, which is to unintentionally trigger an action via eye gaze. We compensate this by adding foot input to clearly communicate a user's intention to perform a certain action, for example, to zoom towards a location currently looked at. Thus, we aim for using eye gaze in a very subtle and unobtrusive way. For this, gaze input is only processed, if appropriate foot input is performed. Without any foot input, a user can freely look around without the concern to unintentionally issue an action. In general, we use gaze input for pointing tasks such as to indicate where to zoom in and for gaze-directed panning based on the distance to the screen center (e.g., see [7, 11]).

Secondly, if gaze is used for panning operations, we have to take into account that fast panning motions via gaze control should be avoided. This is due to the circumstance that eye gaze assumes a double role for orientating oneself in a scene and for controlling the pan direction and speed.

Figure 2: *Three pedals* setup.

Figure 3: *Foot-rocker* device.

Figure 4: *Foot-joystick* & *Foot-rocker* setup

As a result, this may cause disorientation and motion sickness [7]. We propose several approaches to counteract this with foot input for the individual prototypes. This includes controlling the panning speed via foot input to give the user a higher control and to be able to quickly stop in case he/she feels lost. As an alternative, we also have one prototype in which panning is entirely performed via foot controls. More details about how the different interaction tasks are mapped to gaze and are described in the next section when introducing the individual design prototypes for gaze-supported foot interaction.

To investigate suitable combinations of gaze and foot input, we aimed for a diversity of foot input devices, for which we pursued the following design goals:

Comfort: Natural, precise and effortless foot input

Unobtrusiveness: No need for attaching additional equipment to the user's foot or leg

Robustness: The foot device should be sufficiently strong to withstand the weight and force of a human foot and leg

Precision & Customization: High-resolution input for precise movement control and highly adjustable configuration for hardware and software parameters

Intentions: Specific foot-based starting conditions (e.g., a minimum amount a foot pedal has to be pressed down) to avoid involuntarily issuing an action

Prototype Setup

For our software prototype, we use *Google Earth* as a popular representative of a zoomable information space. For tracking gaze data, we use a Tobii TX300. This is a table-mounted, high frequency (up to 300 Hz) binocular eye tracker attached to a removable 23" HD monitor. For foot input, we use three different input devices that we have combined in different ways. First, we briefly describe the three foot input devices and then discuss the respective input combinations for our three prototypes.

Foot pedals. On the one hand, we use the Fanatec CSR Elite[1] pedals as imitation of common car pedals (see Figure 2). The aluminum pedals are robust and offer various configuration possibilities, for example, the angle, order, orientation, as well as horizontal and vertical position of pedals can be adjusted. The setup contains one pressure-sensitive pedal (originally brake) and two pedals depending on angular displacement (originally gas and clutch). All pedals deliver values with a maximum resolution of 10 bits.

Foot joystick and Foot rocker. As an alternative to common car pedals, we designed and manufactured two custom-made foot input devices: *Foot-rocker* (see Figure 3) and *Foot-joystick* (see Figure 4) with two axes. The *Foot-rocker* is a pedal manufactured with a fixed centered single axis that can be tilted about 20° forward or backward. The *Foot-joystick* is a ball joint mounted pedal construction containing a cardan joint from a regular gaming-joystick, giving the device two degrees of freedom and a deviation angle of 20° in all directions (see Figure 5). With the help of a spring mechanism both custom pedals will return to their neutral position from any angular displacement after lifting up the foot. All devices can be connected to a computer via USB using DirectInput. Finally, please note that the devices shown in Figure 3 and Figure 4 are still prototypes and have not yet been optimized in size for a more convenient use.

For our demo, the user sits in front of the eye tracker at a

[1]For further information see http://eu.fanatec.com/

Figure 5: *Foot-joystick*: Two-axes foot-based tilting.

Figure 6: *Foot-rocker*: Two-directional foot pedal.

standard desk with the respective foot device underneath it. The foot devices are not fixed to the ground and can be placed according to personal preferences. We distinguish three particular setup configurations for further investigations that are described in the following.

Prototype 1: **Three pedals**
For the first prototype, we use three foot pedals as they are common for controlling a car. With this similarity, we anticipate that the interaction should be easy to learn and use. Based on our previous description of the Fanatec CSR Elite pedals, the prototype includes one pressure-sensitive pedal that is installed on the left in a standing position (attached from below) while the others are in a hanging position (see Figure 2). A user can pan into the direction, he/she is currently looking at by pressing the left pedal for activation and to increase panning speed. Following promising gaze-based panning approaches [7, 11], this means that a target that is currently looked at will intuitively move toward the screen center. Similarly, a user can zoom towards an area currently looked at by pushing the middle pedal to zoom in and the right pedal to zoom out (see [8]). However, this setup has the disadvantage that the mapping of zooming directions is inconsistent: both zoom pedals have the same operation direction (i.e., a pedal has to be pushed forward), but are associated to contrary virtual movement directions (i.e., zoom in and out). This may confuse users and impede a fluent interaction.

Prototype 2: **Single pedal & Foot-rocker**
To allow for a more intuitive mapping of zooming directions, we combine a single foot pedal (from Fanatec CSR Elite) with a two-directional foot pedal (the *Foot-rocker*) for our second prototype. Analogous to *Prototype 1*, the single pedal is used to control the panning speed towards the current gaze position. Zoom in/out can be performed by tilting the *Foot-rocker* forward/backward. Again the user can indicate where to zoom in/out by looking at a respective location.

Prototype 3: **Foot-joystick & Foot-rocker**
Both *Prototype 1* and *2* use an approach for which the user controls the panning direction via eye gaze. However, fast gaze-based panning across large distances should be avoided as this may lead to disorientation and nausea [8]. To address this issue, we combine both *Foot-rocker* and *Foot-joystick* for our third prototype. The *Foot-joystick* allows for controlling the pan direction in all directions without the need for additional gaze input. Analogous to *Prototype 2*, the *Foot-rocker* is used for controlling the zooming speed, whereby the zooming pivot is based on what the user is currently looking at. This setting enables users to intuitively zoom towards viewed targets and to quickly pan via foot input.

Discussion and Outlook

In this workshop paper, we could only give a first impression of our ongoing investigation of gaze-supported foot input. The prototypes are already implemented and ready for further evaluations. Thus, as a next step we plan to thoroughly evaluate the presented prototypes to learn more about the efficiency and usability of panning and zooming with the particular input combinations. In this context, we are also interested in finding out how users assess these novel post-WIMP controls.

As an outlook, there are several issues that are interesting for further investigation. First, how can gaze-supported foot input be used for additional interaction tasks, such as rotating the view? This would, for example, allow to use these novel controls to ease working with 3D modeling tools or with virtual environments in general. Secondly, it might be worth considering gaze input as a universal pointing modality for both primary and secondary interaction tasks. For this, we could easily distinguish different interaction modes depending on the type of received input (manual or foot). For example, while quick object selections could be performed by looking at an item and pressing a keyboard button, pan and zoom could still be performed by looking at a region of interest and pressing a foot pedal. However, this would have the disadvantage that both primary and secondary tasks could not be performed simultaneously anymore. Finally, further investigations are required to find out how gaze-supported foot interaction could be applied in different application areas, such as multimodal game controls or in medicine. For example, gaze-supported foot interaction could aid surgeons to keep their hands free or in place while operating secondary equipment with gaze and feet.

Conclusion

The multimodal combination of gaze data and foot input to support secondary interaction tasks is very promising. By using gaze-supported foot interaction we have introduced a seamless approach for exploring zoomable information spaces. In our initial investigations, we have developed novel ways for pan and zoom control benefiting from implicit gaze input with explicit foot controls. This enables users to perform secondary tasks in a non-fatiguing way without interrupting manual input. Gaze-supported foot interaction offers many possibilities for the integration in different application fields, but requires further investigations for effortless and efficient gaze- and foot-supported controls.

References

[1] Adams, N., Witkowski, M., and Spence, R. The inspection of very large images by eye-gaze control. In *Proc. of AVI '08*, ACM (2008), 111–118.

[2] Daiber, F., Schöning, J., and Krüger, A. Whole body interaction with geospatial data. In *Smart Graphics*, vol. 5531 of *Lect. Notes in CS*, Springer (2009), 81–92.

[3] Istance, H., Bates, R., Hyrskykari, A., and Vickers, S. Snap clutch, a moded approach to solving the midas touch problem. In *Proc. of ETRA '08*, ACM (2008), 221–228.

[4] Pakkanen, T., and Raisamo, R. Appropriateness of foot interaction for non-accurate spatial tasks. In *Proc. of CHI EA '04*, ACM (2004), 1123–1126.

[5] Pearson, G., and Weiser, M. Of moles and men: the design of foot controls for workstations. *SIGCHI Bull. 17*, 4 (Apr. 1986), 333–339.

[6] Schöning, J., Daiber, F., Krüger, A., and Rohs, M. Using hands and feet to navigate and manipulate spatial data. In *CHI EA '09*, ACM (2009), 4663–4668.

[7] Stellmach, S., and Dachselt, R. Investigating gaze-supported multimodal pan and zoom. In *Proc. of ETRA '12*, ACM (2012), 357–360.

[8] Stellmach, S., Stober, S., Nürnberger, A., and Dachselt, R. Designing gaze-supported multimodal interactions for the exploration of large image collections. In *Proc. of NGCA '11*, ACM (2011), 1:1–1:8.

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

[10] Vertegaal, R. A fitts law comparison of eye tracking and manual input in the selection of visual targets. In *Proc. of ICMI '08*, ACM (2008), 241–248.

[11] Zhu, D., Gedeon, T., and Taylor, K. "moving to the centre": A gaze-driven remote camera control for teleoperation. *Interact. Comput. 23*, 1 (Jan. 2011), 85–95.