H. Reiterer & O. Deussen (Hrsg.): Mensch & Computer 2012
München: Oldenbourg Verlag, 2012, S. 303-312

Investigating Freehand Pan and Zoom

Sophie Stellmach[1], Markus Jüttner[2], Christian Nywelt[2], Jens Schneider[2],
Raimund Dachselt[1]

Interactive Media Lab Dresden, Technische Universität Dresden[1]; Faculty of Computer Science, Otto-von-Guericke Universität Magdeburg[2]

Abstract

The availability of low-cost and flexible tracking systems for hand and body movements is increasing. With this, more thorough investigations for more natural and efficient physical interaction styles are required which take particular limitations of such systems into account, such as the limited ability to track individual fingers. To contribute to this, we describe an investigation of basic hand gestures for the exploration of large information spaces. A set of four pan-and-zoom alternatives using two-handed gestural controls have been implemented and compared using *Google Earth* as an example. For this we conducted a small-scaled formative user study with nine participants to fundamentally assess users' acceptance and the qualification of these freehand gestures for pan-and-zoom operations. As a result, a simple forward and backward hand movement for zooming and a joystick metaphor for panning yielded in the best overall results. Especially the seamless integration of continuous pan and zoom was positively highlighted by participants.

1 Introduction

For the exploration of large information spaces, such as giga-pixel images or geographical information systems (GIS), pan and zoom are essential means of navigation. In addition, such information is often displayed on large-sized distant screens for which hand and body gestures may allow a flexible interaction with. For this purpose, we want to investigate how different hand and arm gestures could be used for the exploration of large information spaces. Several challenges arise for the design of freehand gestures, such as increased physical effort leading to higher fatigue and higher mental effort due to the coordination and memorization of gestures. In this context, Hinckley et al. (1994) report that two-handed interaction may improve efficiency and decrease disorientation compared.

Reliable, flexible, and inexpensive tracking systems to detect hand and body movements are widely available nowadays, such as the *Microsoft Kinect*[1]. In that, they allow for a seamless

[1] Microsoft Kinect Official Web Site. [http://www.xbox.com/kinect] Last accessed June, 2012

and unobtrusive interaction as users have to wear neither additional markers nor data gloves. But particular limitations of such low-cost tracking systems have to be considered for the design of practical interaction techniques, such as the limited ability to detect the movement of fingers. This promotes a deeper investigation of user-friendly and yet efficient low-cost freehand gestures for basic tasks such as the exploration of a virtual information space.

In this paper, we describe and evaluate different freehand gestures for the exploration of large information spaces. With this, we also want to lay a foundation for on-going research on multimodal input for a more natural and efficient interaction in diverse contexts, especially for the interaction with distant displays or multi-display setups. In this context, hand gestures could, for example, be well combined with additional modalities such as gaze input to interact with large-sized screens (e.g., (Koons et al. 1993, Stellmach et al. 2011, Yoo et al. 2010)). For our investigations we used *Google Earth*[2] as an example information space. We derived three pan and three zoom gestures from which four combinations have been tested in a small-scaled formative user study to explore their potential usability in terms of how well user's could cope with them to navigate in Google Earth.

The remaining paper is structured as follows. First, we discuss related work on how freehand gestures have been previously used for panning and zooming. In Section 3, we discuss design considerations and our resulting concepts for freehand pan-and-zoom techniques. Based on this, four selected freehand pan-and-zoom combinations have been tested in a user study that is described in Section 4. We conclude this paper with a discussion of the obtained results and an outlook to continued research on freehand gestures.

2 Related Work

The exploration of virtual data is a fundamental task in various application domains. In this context, hand gestures may allow for natural and yet efficient interaction with distant large-sized displays. In particular, two handed input may improve performance even if tasks are executed sequentially (Buxton and Myers 1986, Gribnau and Hennessey 1998). Furthermore, Mine (1995) points out that two aspects have to be specified to steer (navigate) in a virtual space: direction and velocity of movement. As an example, the distance between hand and body can be used to gradually increase or decrease the velocity. The direction can be specified by a hand pointing. Both may, however, lead to a higher fatigue for the arm (Mine 1995, Bowman et al. 2005). Alternatives for speed control include moving the hand away from a user-defined null position (instead of the user’s body as a reference), tilt the hand, or perform a stop hand gesture (i.e., hold up the hand) to stop moving (e.g., Franke et al. 2010).

Several hand-based input techniques exist for steering in virtual environments. With the “scene in hand” technique (Ware and Osborne 1990), a user can control the virtual camera as if it would be positioned in his/her hand. This approach has been enhanced to the “world in

[2] Google Earth Official Web Site: [http://earth.google.com] Last accessed June, 2012.

miniature" and "grabbing the air" techniques (Mapes and Moshell 1995, Stoakley et al. 1995, Bowman et al. 2005). The first describes a metaphor where the user holds a miniature representation in his/her hand to adapt camera parameters. Thus, moving the hands apart leads to a magnified view at the scene (i.e., zoom in). The latter, the "grabbing the air" technique, lets the user literally grab the world around him/her and pull him-/herself through the virtual scene in any direction using 3D pull hand gestures. Following this, Yoo et al. (2010) present push-and-pull hand gestures to zoom and a combination with gaze/head input to pan. They indicate a high potential for more attentive and immersive interaction with large-scaled displays compared to traditional input devices such as mouse and keyboard.

So far, the wide application of hand gestures in an everyday context has been hindered by the high cost and complexity of tracking systems. In addition, tracking hand gestures has often been obtrusive, because additional markers had to be attached to a user's hands or (data) gloves had to be worn. With low-cost and less obtrusive tracking alternatives such as the Microsoft Kinect, hand-based input can be explored more broadly. For example, Boulos et al. (2011) present several hand gestures for navigating in Google Earth using such a system. However, their techniques require additional finger gestures and special gestures for mode changes which prevent a fast simultaneous interaction (Stannus et al. 2011). Also, Schlattmann et al. (2009) use finger gestures for which they evaluate the hand position and orientation, as well as the orientation of individual fingers. This allows pointing in a direction to which the virtual camera should move to. Similarly, Nancel et al. (2011) use the primary hand for pointing and for indicating the target direction, while different modalities are compared for zooming via the secondary hand. This includes a mouse scroll-wheel, a mobile touchscreen, and freehand gestures. However, they indicate that the freehand gestures were less accurate and efficient compared to the handhelds.

In a nutshell, several approaches for hand-based input for steering in virtual environments have been presented over the years. However, it remains somewhat unclear how simple freehand gestures have to be designed to allow for an efficient and yet user-friendly pan and zoom interaction. While freehand gestures have the advantage of not needing to hold an additional device and of a high flexibility to be applied in various application contexts, they are often not suitable for precise pointing tasks.

3 Concept for Freehand Pan and Zoom

We elaborated a set of basic freehand pan-and-zoom techniques (see Figure 1 for an overview) with the aim of providing user-friendly and yet affordable techniques using a low-cost tracking system, such as the Microsoft Kinect. After briefly discussing some basic design considerations, the pan and zoom techniques that have been selected for further investigation in a user study are described.

One design aspect to consider is whether individual interaction modes should be available separately. While dissociating individual pan and zoom modes may allow higher control for users, time-consuming changes between different navigation modes could quickly become

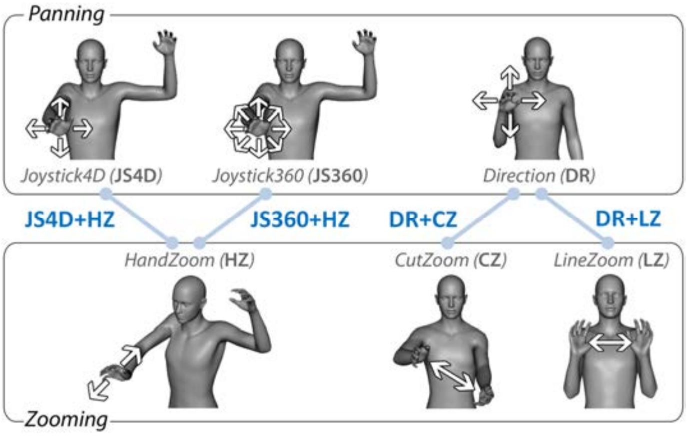

Figure 1: Overview of the proposed basic freehand pan and zoom techniques.

tedious. In contrast, simultaneously issuing pan and zoom commands could lead to more complex hand gestures and could be more confusing to perform.

Furthermore, we considered how to ease the effort for learning new pan-and-zoom hand gestures. For this, we decided to take advantage of commonly known multitouch pinch-to-zoom techniques, because of the high popularity of multitouch devices. Thus, we decided to consider this principle for the design of our freehand zooming gestures (see Section 3.3).

Finally, freehand gestures are often considered error-prone for quickly and reliably performing a desired task. Thus, careful attention has to be paid to possibilities to avoid unintentionally issuing commands. In this context, the secondary hand may indicate a user's readiness to pan or zoom and thus acts as a delimiter for separating gestural input from normal motion. However, if holding up the secondary hand is applied as such a delimiter, it is not possible to use bimanual pan and zoom gestures. Thus, an additional design aspect is the use of one- vs. two-handed gestures in connection with a suitable *readiness indicator*.

3.1 Freehand Panning

We elaborated two types of panning hand gestures: the joystick gesture (**JS**) and the direction gesture (**DR**) (see Figure 1). For **JS**, both hands are required for panning. First, the primary hand is lifted to a position that is later used as reference (null) position. Then the secondary hand is lifted up to the shoulders to indicate a user's readiness to pan. Therefore, if the secondary hand is not leveling the shoulder, no action will be performed and the user can freely move without unintentionally issuing a command. So, while the secondary hand is lifted, a movement of the primary hand relative to the reference position is interpreted as pan. For this, we distinguish two further variations: restricting the panning to four discrete directions (i.e., up, left, down, and right) (**JS4D**) and additionally allowing diagonal directions (**JS360**). To restrict panning directions for **JS4D** may offer a higher control for issuing a desired command and with that decrease incorrectly issued commands and the required mental effort. On the other hand, **JS360** allows for a smooth transition between different steering (panning) directions.

For **DR**, only one hand is required for panning. The relative change in hand position with respect to the same-sided shoulder is used for panning. If lifting the second hand, no actions will be issued. Thus, the secondary hand acts as a *safety brake* for **DR**.

3.2 Freehand Zooming

We have investigated three different alternatives for freehand zooming: **Hand-Zoom** (**HZ**), **Cut-Zoom** (**CZ**), and **Line-Zoom** (**LZ**) (see Figure 1). First, **HZ** works similar to **JS**. The lifted secondary hand indicates the user's readiness; the primary hand issues the actual commands. Thus, based on the primary hand's initial position (when lifting the secondary hand), a user can zoom in by moving the primary hand closer to the display and zoom out by moving it away from it.

The other two techniques **CZ** and **LZ** resulted from considerations about how to apply common multitouch pinch-to-zoom techniques to freehand gestures. Thus, both techniques are two-handed gestural controls and with that alternative ways how to indicate a user’s readiness (i.e., delimiters) have to be defined. For Cut-Zoom (**CZ**), we decided for a predefined starting position by joining both hands to indicate the intention to zoom. Moving the hands apart results in a continuous zoom (see Figure 1). Since the movement always starts with closed hands, a way to distinguish between zooming in and out has to be specified. For this, we decided for a simple differentiation of which hand is above the other one. This means, if the right hand is highest, the view is zoomed in or analogous zoomed out for the left hand. The distance between both hands determines the zooming speed. Thus, if joining the hands again, the zooming will stop.

For Line-Zoom (**LZ**), both hands have to be at shoulder level to indicate readiness. If moving both hands apart, the view is zoomed in and if moved together, the view is zoomed out. Thus, it is essentially similar to the “world in miniature” metaphor (Stoakley et al. 1995). The movement stops, as soon as one of the hands leaves the shoulder level.

4 User Study

As a next step, we wanted to further explore the potential of the described techniques for the interaction with distant displays. First, we conducted an initial pre-study to decide on promising combinations of the described pan and zoom techniques for further investigation. We selected four alternatives that are listed in Table 1 (also see Figure 1) that were then further investigated in a formative user study. This study mainly aimed at finding out about the fundamental suitability of these techniques in terms of how well users would cope and be satisfied with them to perform desired pan-and-zoom operations. We briefly describe the main differences between these four combinations in the following.

	x-handed Zoom	Restricted Panning	Pan & Zoom Modes	Readiness Indicator
JS4D + HZ	One	Yes	Integrated	Continuous
JS360 + HZ	One	No	Integrated	Continuous
DR + CZ	Two	Yes	Distinct	Discrete
DR + LZ	Two	Yes	Distinct	Continuous

Table 1: Selected pan and zoom combinations and characteristics for further investigation in our user study.

First, the techniques differ in the number of hands that need to be used to perform a zoom gesture. While **HZ** only requires one *active* hand for zooming (the second hand only acts as passive delimiter), **CZ** and **LZ** both require two. The proposed panning techniques only require one *active* hand. Thus, a seamless integration of the one-handed panning with one-handed zooming techniques is feasible. In contrast, for **DR+CZ** and **DR+LZ** distinct pan and zoom modes have to be distinguished, which may offer higher control for the user. Furthermore, the techniques differ in the way how to indicate a user's readiness (i.e., gesture delimiter). While for **JS4D+HZ** and **JS360+HZ**, the user has to continuously lift the secondary hand, the user only needs to join hands at the beginning for **DR+CZ**. For **DR+LZ**, the hands have to be (continuously) held at shoulder level to zoom. Finally, the panning types differ in the selection of moving direction: four distinct vs. seamless 360° directions.

Participants. Nine participants (3 female, 6 male) volunteered in the within-subject user study, aged from 20 to 25 (Mean (M) = 22.0). None of them had prior experience with free-hand gestures for navigation. However, three participants mentioned that they frequently play games with the Kinect on the Xbox360. Based on a 5-Point-Likert scale from *1–Do not agree at all* to *5–Completely agree*, participants rated several statements about their background. Based on this, they indicated that all are familiar with Google Earth, however, do not use it particularly often (M=3.1, Standard Deviation (SD)=0.9).

Apparatus. For the interaction with a large information space, we use Google Earth as an example. For this purpose, we implemented a Microsoft Windows Forms tool based on C# that uses the Google Earth plug-in. A Microsoft Kinect has been used as a low-cost system to track hand gestures. For the software development for the Kinect we used the *OpenNI*[3] and *PrimeSense Nite*[4] frameworks that we integrated in the Windows Forms tool. Furthermore, a ceiling-mounted projector provided a large projection (approximately 2.5 meters wide and 1.5 meters high). Participants stood about 2 meters away from the projection wall, and the Kinect was positioned below it (see Figure 2).

Procedure. Participants were welcomed, briefly introduced to the study and were asked to answer an initial questionnaire about their demographic background and their familiarity

[3] PrimeSense Nite - Official Web Site: [http://www.primesense.com/Nite] *Last accessed June, 2012.*

[4] OpenNI - Official Web Site: [http://openni.org] *Last accessed June, 2012.*

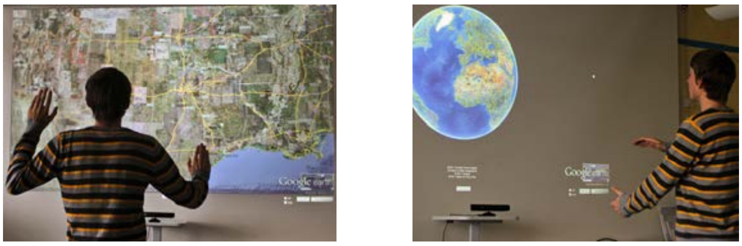

Figure 2: Setup as deployed in the reported user study.

with Google Earth. Then participants tested the four combinations of the described freehand pan-and-zoom techniques in a counter-balanced order (see Figure 1). Each technique could be tested until participants felt sufficiently acquainted with the respective hand gesture. The actual task in the user study was to reach five different cities from the same start position (25N 0' 0", 40W 0' 0" and ca. 16 miles above the ground) one after another in a given order: *Vancouver*, *Sydney*, *Brasilia*, *Gothenburg*, and *Mumbai*. The current target city was marked in Google Earth with a virtual pin icon and was also shown on a little world map printed on a sheet of paper before starting the time measurement. The task completion time for each target was measured individually. A target was considered reached if getting below a preset range (height level) and distance to the target. After a pan-and-zoom combination has been tested, participants were asked to evaluate it in an intermediate questionnaire (see *Section Measures*). After all combinations have been tested, a final questionnaire was handed out. Each participant took approximately 70 minutes on average to complete the study.

Measures. The quantitative measures included task completion times for reaching the five target cities. In addition, we lay a particular high emphasis on user feedback to assess the usability of the individual techniques. For this, an *intermediate questionnaire* has been handed out after the five target cities had been reached with a respective pan-and-zoom technique. The intermediate questionnaire contained two types of questions that were the same for each pan-and-zoom combination:

- **(Q1)** Participants were asked to rate sixteen statements referring to eight usability aspects (two for each, see Figure 4) based on a 5-Point-Likert scale from *1–Do not agree at all* to *5–Completely agree*. This means that respondents specified their level of agreement or disagreement on a discrete symmetric agree-disagree scale for a series of statements. Thus, the range captures the intensity of their feelings for a given item. The eight usability aspects listed in Figure 4 are based on quality factors that describe the effectiveness for travel techniques from Bowman et al. (1997) and comply with a similar pan-and-zoom study from Stellmach and Dachselt (2012) (using gaze input though).
- (**Q2**) Two questions asking for qualitative feedback on what the users particularly liked and disliked about the tested pan-and-zoom techniques.

In the final questionnaire, participants were asked to rate each of the four pan-and-zoom combinations based on a 5-Point-Likert scale with *1–Did not like at all* to *5–Liked it very much* to assess them in contrast.

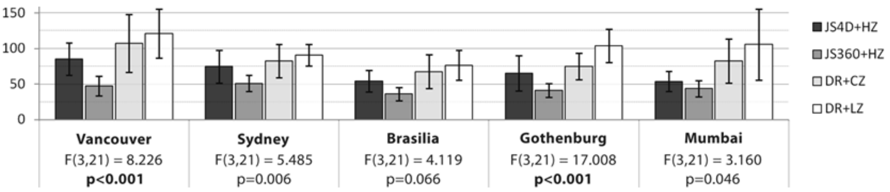

Figure 3: Accumulated mean task completion times in seconds with 95% confidence intervals.

5 Results

Task Completion Times. A repeated-measures ANOVA (Greenhouse-Geisser corrected) with post-hoc sample t-tests (Bonferroni corrected) were used to investigate task completion times (see also Figure 3). Task times did not differ significantly among the five targets for a respective pan-and-zoom technique. However, they differed significantly among the four techniques (F(3,105)=25.55, p<0.001). Using **JS360+HZ** participants were quickest for all five targets. In fact, task times for **JS360+HZ** were significantly faster going to *Vancouver, Sydney* and *Gothenburg* than for **DR+LZ** and **DR+CZ**. In addition, participants were in general faster using **JS4D+HZ** compared to **DR+CZ** and **DR+LZ**, however, slower than using **JS360+HZ**. While participants needed longest with **DR+LZ** for all target cities, no significant differences could be identified between **DR+LZ** and **DR+CZ**.

User Feedback. The answers from the intermediate questionnaires are summarized in Figure 4 according to different usability aspects. In general, user preferences clearly tended towards **JS4D+HZ** and **JS360+HZ** except with respect to *speed* for which all techniques were assessed similar. The combination of **JS360+HZ** received highest ratings among the tested techniques except for *spatial awareness* (i.e., users did not feel disoriented after performing a movement) for which both **JS360+HZ** and **JS4D+HZ** were assessed very positively. In contrast, **DR+CZ** and **DR+LZ** in general received lower ratings. In particular they were assessed as imprecise and not intuitive. In addition, participants often could not accomplish actions as anticipated (cf. Figure 4, *Task-driven use*). However, several participants mentioned that the **CZ** gesture was fun to use for zooming.

After having tested all four gesture combinations, participants were asked to rate which of the zoom and pan techniques they would prefer on the previously described 5-Point-Likert scale. For *zooming*, participants preferred **HZ** (M=4.25, SD=0.83), for which the primary hand had to be moved towards or away from the display to zoom in or out. For *panning*, participants preferred the joystick metaphor and especially highlighted the 360° panning variant **JS360** (M=4.63, SD=0.48).

Finally, participants offered several suggestions on how the tested techniques could be improved. Nearly all participants proposed that it should be possible to set the initial reference point for **JS4D** and **JS360** without the need to permanently holding up the secondary hand, as this was tiresome after a while. In this context, the suggested integration of these techniques with other modalities, such as speech or a handheld, was found interesting. Further-

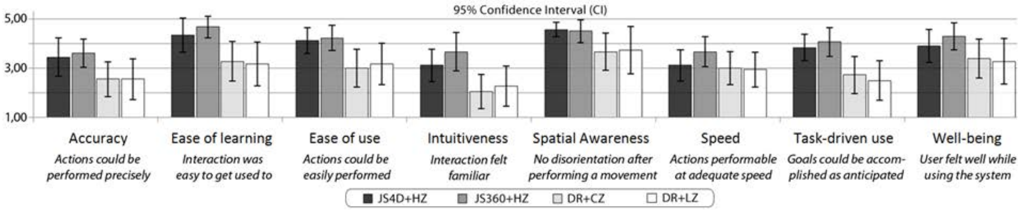

Figure 4: Quantitative user feedback based on 5-Point-Likert scales (from 1-very low to 5-very high).

more, although **CZ** resulted in slow task completion times, several participants mentioned that this was the technique that was most fun.

6 Discussion & Conclusion

The results indicate that the interaction with the pan and zoom combination **JS360+HZ** offers high potential for an efficient and yet user-friendly navigation as it achieved the overall best task times and was assessed positively by participants. In contrast, the worse results for **DR+CZ** and **DR+LZ** indicate that a continuous transition from panning to zooming is desirable and that an imitation of common touch pan and zoom gestures may not be beneficial after all. Thus, the advantage of higher control by clearly dissociating between pan and zoom could not prevail against the discomfort in persistently switching between modes. However, as some participants indicated the fun potential of the **CZ** zoom gestures, it could be interesting to investigate how a combination with an alternative panning modality would be assessed, such as a gaze-supported control (e.g., (Stellmach et al. 2012)). This would allow for two non-conflicting simultaneous input channels avoiding mode switches.

In this paper, we investigated several basic freehand gestures for panning and zooming in large information spaces. As an example, we tested and evaluated four combinations of pan and zoom techniques in Google Earth using a Microsoft Kinect with nine participants in a small-scaled formative user study. After all, a simple forward and backward hand movement for zooming and a hand panning based on a joystick metaphor resulted in the best overall results. Our main aim was to investigate the usability of freehand gestures that can easily be applied in various user contexts. With that, we also aimed at providing a foundation for ongoing work for multimodal interaction with distant displays incorporating hand gestures.

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

Boulos, M. N., Blanchard, B. J., 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):45.

Bowman, D.A., Koller, D. & Hodges, L.F. (1997). Travel in Immersive Virtual Environments: An Evaluation of Viewpoint Motion Control Techniques. In Proc. VRAIS '97, IEEE, 45-52.

Bowman, D.A., Kruijff, E., LaViola, J.J. & Poupyrev, I. (2005). 3D User Interfaces – Theory and Practice. Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA.

Buxton, W. & Myers, B. (1986). A study in two-handed input. In Proc. of CHI '86, New York, NY, USA, ACM, 321-326.

Franke, R., Koch, M., Stellmach, S. & Dachselt, R. (2010). Intuitives zweihändiges Arbeiten in der virtuellen Realität. GI VR/AR Workshop '10, Shaker Verlag, Aachen, pp. 107-118.

Gribnau, M.W. & Hennessey, J.M. (1998). Comparing single- and two-handed 3D input for a 3D object assembly task. In Proc. of CHI '98, New York, NY, USA, ACM, 233-234.

Hinckley, K., Pausch, R., Goble, J. C. & Kassell, N. F. (1994). A survey of design issues in spatial input. In Proc. of UIST '94, New York, NY, USA, ACM, 213-222.

Koons, D., Sparrell, C. & Thorisson, K. (1993). Integrating simultaneous input from speech, gaze, and hand gestures. In American Association for Artificial Intelligence '93, 257-276.

Mapes, D. & Moshell, J. (1995). A Two-Handed Interface for Object Manipulation in Virtual Environments. Presence: Teleoperators and Virtual Environments, 4(4), 403-416.

Mine, M.R. (1995). Virtual environment interaction techniques. Tech. report, Chapel Hill, NC, USA.

Nancel, M., Wagner, J., Pietriga, E., Chapuis, O. & Mackay, W. (2011). Mid-air pan-and-zoom on wall-sized displays. In Proc. of CHI '11, New York, NY, USA, ACM, 177-186.

Schlattmann, M., Broekelschen, J. & Klein, R. (2009). Real-time bare-hands-tracking for 3D games. In IADIS Intl. Conference Game and Entertainment Technologies (GET '09), IADIS Press, 59-66.

Stannus, S., Rolf, D., Lucieer, A. & Chinthammit, W. (2011). Gestural navigation in Google Earth. In Proc. of OzCHI '11, New York, NY, USA, ACM, 269-272.

Stellmach, S. & Dachselt, R. (2012). Investigating Gaze-supported Multimodal Pan and Zoom. In Proc. of ETRA '12, New York, NY, USA, ACM, 357-360.

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

Stoakley, R., Conway, M. & Pausch, R. (1995). Virtual reality on a WIM: interactive worlds in miniature. In Proc. of CHI '95, New York, NY, USA, ACM, 265-272.

Ware, C. & Osborne, S. (1990) Exploration and virtual camera control in virtual three dimensional environments. In Proc of the 1990 symposium on Interactive 3D graphics, ACM, 175-183.

Yoo, B., Han, J.-J., Choi, C., Yi, K., Suh, S., Park, D. & Kim, C. (2010). 3D user interface combining gaze and hand gestures for large-scale display. In Proc. of CHI EA '10, ACM, 3709-3714.

Contact Information

Prof. Raimund Dachselt (Technische Universität Dresden)
Telefon: (+49) 351 / 46338507
E-Mail: dachselt@acm.org