Comparing Gaze-based and Manual Interaction in a Fast-paced Gaming Task in Virtual Reality

Felix Hülsmann, Timo Dankert, Thies Pfeiffer

Artificial Intelligence Group, Faculty of Technology, Bielefeld University
Universitätsstraße 25, 33615 Bielefeld, Germany
email: {fhuelsma, tdankert, tpfeiffe}@techfak.uni-bielefeld.de

Abstract: The idea of using gaze as an interaction modality has been put forward by the famous work of Bolt in 1981. In virtual reality (VR), gaze has been used for several means since then: view-dependent optimization of rendering, intelligent information visualization, reference communication in distributed telecommunication settings and object selection. Our own research aims at improving gaze-based interaction methods in general. In this paper, gaze-based interaction is examined in a fast-paced selection task to identify current usability problems of gaze-based interaction and to develop best practices. To this end, an immersive Asteroids-like shooter called Eyesteroids was developed to support a study comparing manual and gaze-based interaction methods. Criteria for the evaluation were interaction performance and user immersion. The results indicate that while both modalities (hand and gaze) work well for the task, manual interaction is easier to use and often more accurate than the implemented gaze-based methods. The reasons are discussed and the best practices as well as options for further improvements of gaze-based interaction methods are presented.

Keywords: Virtual Reality, Interaction, Gaze-based Interaction, Eyetracking, Immersion

1 Introduction

Enabling a computer system to read our wishes in our eyes is a fantastic dream which has, inter alia, been visionized by Bolt [Bol81]. Since then, this dream has started to become true, step by step.

In theory, the advantages of gaze-based interaction are clear: the eyes are the fastest modality we have, eye movements even preceed other modalities, such as speech or pointing gestures, and eyes reveal information about their owners' visual attention, which can be used to identify wishes even before an explicit instruction has been given. There are also situations where the hands are already engaged in a task and thus additional modalities are required, e.g., to interact with a digital assistance system. Examples of this are medical surgeries, which could greatly benefit from augmented information displays guiding the tools of the surgeon, or assemblages, e.g., in car construction or repair. Today, gaze-based interaction is already a viable surrogate for manual interaction for handicapped persons. The most common application in this context is gaze-typing.

The current methods for gaze-based interaction, however, also entail several disadvantages, which have to be overcome: the required gear is obtrusive and, as gaze-based interaction is underexplored compared to manual interaction, there is a lack of experience guiding gaze-based interaction design.

In a scientific niche, gaze-based interaction has been tested and analyzed for many years targeting desktop applications [CH87]. Research on using gaze in 3D applications when immersed in virtual reality or when interacting with the real world in augmented reality, however, is less common. Here, besides pure benefits regarding interaction performance, versatile gaze-based interaction could also improve immersion by replacing interaction devices that are complex to handle. The section on related work will report the most prominent approaches.

With the current advent of consumer 3D technology, such as 3D television and 3D tracking (Sony PlayStation Move, Microsoft Kinect), and similar developments regarding gaze tracking (miniaturization, open source tracking software), gaze-based interaction in 3D requires an increased attention in interaction design.

In this paper, sighting and selection are addressed as interaction tasks commonly considered for gaze-based interaction. Therefore, two studies comparing manual and gaze-based interaction in a fast-paced gaming task are presented. As testbed, an immersive game called "Eyesteroids" was developed, which is inspired by Asteroids published by Atari in 1979. In this game, the user is situated in outer space and has to fight moving space ships. The space ships have a weak spot which the user is challenged to hit. In Eyesteroids, **sighting** means aiming at a hostile object, whereas **selection** means triggering the shot.

The next section presents related work on gaze-based interaction. After that, the Eyesteroids game and the hardware setup are explained in more detail in Section 3. Subsequently, in Section 4, the implemented interaction methods are presented. Section 5 and Section 6 report on two studies conducted based on Eyesteroids. Section 7 concludes with a discussion of lessons learned and best practices as well as opportunities for future work.

2 Related Work

Concerning the usability of gaze-based interaction methods, there is little clear empirical evidence: Tanriverdi and Jacob compared gaze-based interaction with a hand-pointing approach [TJ00]. They showed, that gaze-based interaction was faster than pointing with the hands, especially for distant objects. They also pointed out that users were equally satisfied with both types of interaction. On the other hand, Cournia, Smith and Duchowski[CSD03] did not find convincing results for a gaze-based interface compared to hand-based interaction methods: For distant objects, gaze-based selection was slower than a hand-based one. Concerning the ease of use, Ohno and Hammoud described various problems in the use of eye tracking devices [OH08]. They especially pointed out the discomfort of those devices, gaze recognition problems through eyeglasses or contact lenses and the need to avoid the *Midas Touch Problem*. Aiming with gaze in first person shooter games has been presented and evaluated by Isokoski and Martin [IM06]. The aiming has been done by gaze; for other interaction modes they used keyboard and mouse. The results were not fully satisfying. Smith and Graham also presented gaze-based user interfaces for different video games [SG06]. Here, a large number of users felt more immersed in the game's world using the gaze-based interaction than using the mouse. However, the performance and the eyetracker's acceptance in the user's point of view were dependent on the type of game the user was playing.

In our former work, a lightweight head-mounted eye tracking system was integrated in a CAVE-like virtual reality set-up and evaluated afterwards [Pfe08]. Also, algorithms for interaction in VR were developed and evaluated [PLW09]. In this work, we build on these information and compare the gaze-based interaction with other interaction methods.

3 Eyesteroids

3.1 The Eyesteroids Game

Figure 1: Participant in the Eyesteroids scenario

Our test scenario is based upon Asteroids, a videogame published by Atari in 1979. The user is situated in a virtual space where he has to aim at circular targets and shoot them as accurately as possible. The targets are moving on a plane. The user plays in a first-person perspective. A sonification of the shoots and a red light ray miming a laser ray are added to improve immersion. During the whole game, the user is not limited in his movements; he is allowed to walk freely around in the CAVE. Figure 1 shows a participant playing Eyesteroids.

3.2 The Equipment

Eyesteroids is implemented for the use in a three sided CAVE with a polarized light stereo projection and an ART tracking system. The viewpoint is relative to the position and orientation of the user's head.

Figure 2: Nintendo Wii Remote extended by a pistol grip

As pointing-based device, the cordless **Nintendo Wii Remote** with eleven buttons, an accelerometer and 2D-tracking by IR (which was not used in this study) is used (see figure 3a). It is equipped with additional markers in order to track its position and orientation with the ART system. Furthermore it is extended by a pistol grip (see Figure 3b).

For gaze tracking, a wired **Arrington Research eye-tracking system**: ViewPoint PC-60 BS007 is used, which is mounted on polarized glasses for the stereo view (see figure 3). The eye tracker's technical data can be found in Table 1. The developed software considers orientation of the eyes, pupil size and ratio between height and width of the pupil due to cameras which record the eyes from below. Optical markers of the ART system are mounted

Angular Accuracy	0.25° − 1.0°
Angular Precision	0.15°
Temporal resolution	30/60Hz
Optical resolution	640x480/320x240 Pixel

Table 1: Technical data of the eye-tracker

on the eye tracker to determine the user's position and orientation, which is also used to adapt the point of view.

4 Interaction Modes

Figure 3: View-Point PC-60 BS007

Interaction methods After creating the scenario, a set of six interaction methods was developed to compare the use of gaze-based methods with manual interaction methods like pointing using the Nintendo Wii Remote. Some of these methods were multimodal (e.g. selecting with the eyes, sighting with the Wii Remote), some were unimodal. The goal was to design the interaction methods as intuitive and easy to use as possible.

4.1 Sighting

In Eyesteroids, sighting means aiming at the hostile object. Three different methods of sighting targets seemed reasonable. These methods differ in some ways, but they are all based on ray-casting, because this method provides a good performance and is suitable for novice users [RDS+10].

Gaze The user focuses the target he wants to shoot at with his dominant eye. The viewing direction is recognized by using the eyetracker.

Lightgun The modified Nintendo Wii Remote is used for sighting. To increase immersion, a prop has been used to give it the shape of a Lightgun. In this mode, there is the possibility of using an ironsight.

Limited Lightgun This mode is similar to the previous one: The user operates the modified Nintendo Wii Remote, but this time he is not allowed to use the ironsight: He has to shoot from the hip. This method is used to find out, if it is the ironsight that has an impact on the interaction performance in terms of speed and accuracy.

4.2 Selection

Selection means triggering a shot. For our first study, two kinds of selection methods were developed: Blinking with one eye and pulling the pistol's trigger. After evaluating, it seemed necessary to improve the blinking by developing an additional method: Blinking with both eyes.

Blink (one eye) The user can trigger the shoot by closing his non-dominant eye. The dominant eye has to remain open. There are two reasons for using this combination:
On the one hand, the user does not lose immersion due to closing his eyes. He also gets feedback, as the shoot is visualized by a light ray. On the other hand, there are technical reasons for leaving one eye open: The intended blink is easier to distinguish from normal eye blinking and (if the sighting method "Gaze" is selected), the user can still aim with his dominant eye.

Blink (both eyes) In this mode, the user triggers the shoot by blinking with both eyes: The eyes are closed for approximately one second to distinguish this intentional blink from normal blinks. This method has the disadvantage of decreasing the users immersion due to the missing visual feedback.

Pistol trigger The pistol trigger mode allows the user to shoot by pulling the pistol's trigger. Feedback is provided by the light ray (see figure 1).

5 First Study

The above presented sighting and shooting methods were evaluated in two studies. In the first one, the combinations listed in table 2 were tested.

ID	**Sighting**	**Selection**
g/b	*gaze*	*blink (one eye)*
g/p	gaze	pistol trigger
l/b	lightgun	blink (one eye)
l/p	*lightgun*	*pistol trigger*
ll/b	limited lightgun	blink (one eye)
ll/p	*limited lightgun*	*pistol trigger*

Table 2: Combinations for the first study, unimodal combinations are highlighted in italics.

Twenty participants ranging in age from 17 to 49 years (median: 22) took part. People with and without experience with virtual worlds were tested and the number of participants with a dominant left or right eye was balanced. Before starting the study, the participants were introduced to the scenario and were told to shoot at the moving targets as precisely as possible. Second to that, all possible selection and sighting methods were explained.

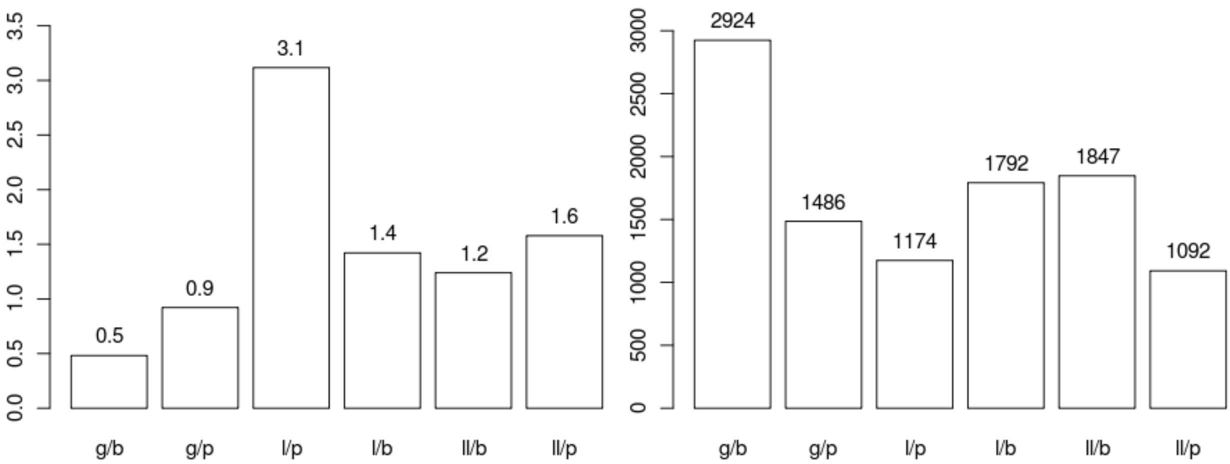
The study consisted of six cycles, distinguished by the use of different combinations of interaction methods. The sequence was chosen at random. At the beginning of each cycle, the eyetracker was recalibrated. Then the explanation of the interaction methods chosen for the current cycle was repeated. Before starting the cycle, the participants had to train the current interaction methods by shooting at nine static targets to be sure that each participant understood the function of the current interaction mode. After the training, the evaluation mode began: the participants had to shoot the moving targets. At the end of each cycle, they had to respond to a set of questions to evaluate the subjective factors influencing their immersion (joy of use, exhaustion, user's favourite interaction mode). The objective factors (accuracy, time span between shoots) were evaluated automatically.

5.1 Results

5.1.1 Objective Factors

The objective factors were destinguished between **accuracy** and **speed**. Speed was measured as the time span between two shoots. The participant's shooting accuracy was measured as the ratio between succeeded and failed shoots. Higher values stand for higher accuracy. Mode `l/p` turned out to be the most accurate mode with an accuracy measure of 3.1 (mean). Mode `g/b` was the most inaccurate one with an accuracy measure of 0.5 (mean). All methods with sighting using the lightgun turned out to be more accurate than those using the eye tracker. See figure 4.a for more detailed information.

Concerning the time span between two shoots, mode `ll/p` proved to be the one with the smallest distance between two shoots (1.1 seconds at median), followed by `l/p`. Mode `g/b` had the worst results with 2.9 seconds (median). Further information can be found in Figure 4.b.



(a) Accuracy (the higher the better)   (b) Time between two shoots (the lower the better)

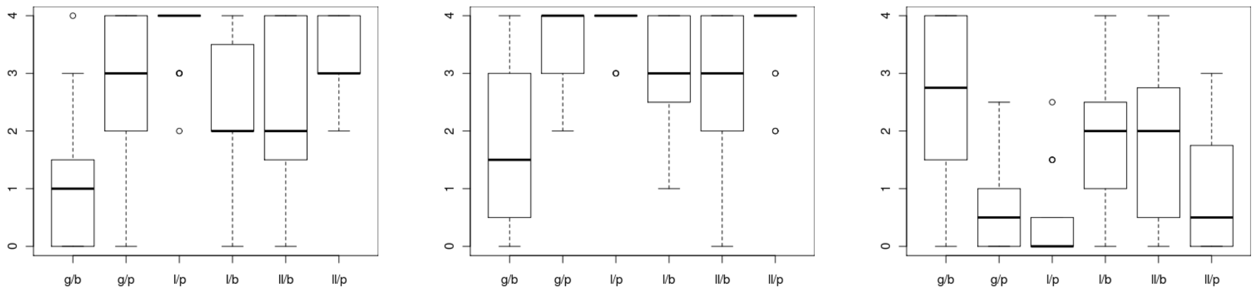
Figure 4: Evaluation of the objective factors in study 1. The axis labels correspond to the IDs in table 2.

5.1.2 Subjective Factors

Three kinds of subjective factors were evaluated: **Easiness** (Figure 5.a), **Joy of use** (Figure 5.b) and **Exhaustion** (Figure 5.c).

After each cycle, the participants rated the easiness of the current interaction method with values from 0 (very difficult) to 4 (very easy). Mode `l/p` was the easiest mode in the user's opinion (median: 4). The most difficult mode was mode `g/b` (median: 1.0). The results concerning joy of use were similar, mode `l/p`,`ll/p`,`g/p` got a median of 4. Mode `g/b` had the worst result with a median of 1.5.

Regarding the exhaustion the participants had to choose values from 0 (not at all exhausting) to 4 (very exhausting). Interaction methods which were based on blinking as the shooting trigger turned out to be the most exhausting ones. Nearly all participants complained about the effort necessary for closing one eye while sighting with the other one.



(a) Easiness (the higher the better)

(b) Joy of use (the higher the better)

(c) Exhaustion (the lower the better)

Figure 5: Evaluation of the subjective factors in study 1

The complaints about the one-eyed blinking method motivated an update of the gaze-based selection method to support blinking with both eyes. The advantage over the old method should be the improved ease of use, the disadvantage is the potential loss of immersion due to closing the eyes for a long time span (approx. 1 second) to distinguish conscious blinks from natural ones. To evaluate this updated method, a second study was run, which is presented in the following section.

6 Second Study

The second study was designed similar to the first one, but with only two cycles instead of six (see Table 3).

ID	**Sighting**	**Selection**
g/bb	gaze	blink (both eyes)
l/p	lightgun	pistol trigger

Table 3: Combinations for the second study.

Ten participants, ranging in age from 21 to 29 years (median: 22), took part in this study. All of them had experience with virtual worlds.

6.1 Results

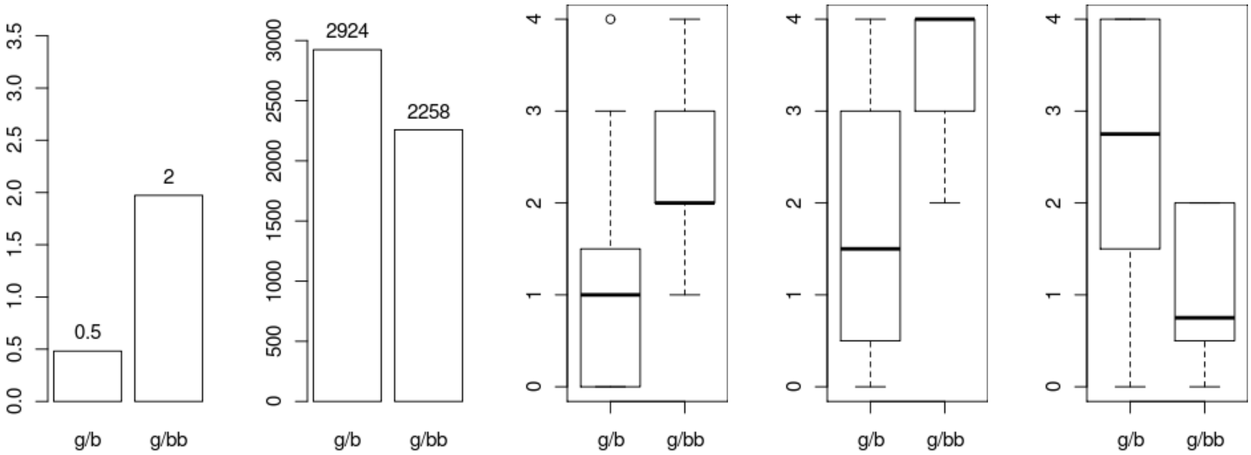
For evaluation, the same criteria as in the first were used. The manual method using the lightgun was used as a baseline to verify the comparability of the two studies. The new results replicated those from the first study and thus testified that.

As expected, the new gaze-based interaction mode g/bb achieved much better results in all categories compared to the unimodal gaze-based approach g/b used in the first study.

6.1.1 Objective Factors

The accuracy of the g/bb method was four times the accuracy achieved with the g/b method in study one (see Figure 6a).

Concerning the time span between two shoots, the g/bb method was about 700 ms faster (2258 ms compared to 2924 ms) than the g/b method (see Figure 6b).



(a) Accuracy (the higher the better) (b) Time between two shoots (the lower the better) (c) Easiness (the higher the better) (d) Joy of use (the higher the better) (e) Exhaustion (the lower the better)

Figure 6: Evaluation of the objective factors in study 2. Left are results from the first study (blink with one eye), on the right are the results from the second study with both eyes.

6.1.2 Subjective Factors

The easiness of the g/bb method was rated one level better than that of the g/b method (see Figure 6c). Participants rated the joy of use of the g/bb method with a median of 4, as opposed to only 1.5 for g/b (see Figure 6d). Also, the g/bb mode was less exhausting than g/b (see Figure 6e).

7 Discussion and Future Work

Our studies show **several advantages of using a manual pointing-based approach** for manual interaction in a fast-paced gaming task in VR. The unimodal manual solutions had a better performance concerning accuracy and speed. In the first study, the accuracy using these modes was nearly six times better than the accuracy with the unimodal gaze-based solution. A reason could be that sighting by aiming with a gun and triggering a shoot using a button is an established interaction method.

Also the time between two shoots was much shorter for the unimodal manual method compared to other methods, which could be explained by the higher certainty of the participants for hitting a target. This hypothesis is supported by the results of the subjective factors: The participants rated the easiness of the manual interaction methods much better

than the one of the gaze-based approach.
Regarding immersion and joy of use, the participants reported prefering the gunlike input device. This may be due to the fact that no adaption is needed for using the manual interaction mode and an appropriate feedback using a projected light ray (see 4.2) is provided.

An explanation for the **gaze-based sighting results** could be the eyetracker itself: It was observed that due to the thick and heavy cable (approx. 1 cm diameter), the participants did not dare to move their head nor their whole body, so that the viewing angles of their eyes often left the best calibrated area. Another explanation for this effect is the danger of drifted eyetracking cameras during the interaction: Some participants scratched their head, which produces a displacement of the tracking cameras and so the calibration becomes inaccurate. Such drift errors often showed up when recalibrating the eyetracking system between the cycles of the study. Also the participants complained about the unfamilar feedback while using the gaze-based approach for sighting: As the user's eyes cannot be the source of the light ray, because of occlusion effects, it was designed to come from the right of the user's head, which was reported as beeing confusing.

In the first study, it was also shown, that all **sighting methods suffered from being combined with a gaze-based selection method**: Concerning the subjective factors, these methods had very poor results. The main reason was the lack of ability of the participants for closing the non-dominant eye while focussing their target with the dominant one. Most of them reported being very exhausted after these cycles. The objective factors (accuracy and speed) also suffered. This could be because the participants were distracted by selecting, so that they were unable to concentrate on sighting at the same time.

As improving gaze-based interaction and developing best practices is the goal of our current work, the unsatisfactory results of the single-eyed gaze-based selection method made us designing the **new unimodal gaze-based interaction method** described in 4.2 as an improvement of the first approach: The goal was to introduce an interaction method which takes advantage of the fact, that an unimodal gaze-based interaction allows the user to keep his hands free for other activities (see Section 1). The new interaction method was evaluated in a second study which was designed to be similar to the first one. This new approach was accepted by the participants concerning joy of use and easiness. Also exhaustion decreased, because of the possibility to close both eyes, which is more natural than leaving one eye open. On the other hand, nine of ten participants still preferred the manual approach over the gaze-based approach when having to choose their favorite interaction mode. This could be explained with a loss of immersion while having the eyes closed: Many participants complained getting no visual feedback except from the score counter after hitting an object.

To sum it up, the results show that the manual interaction method is easy to explain and to use and it does hardly disturb the user's immersion. It also provides a great accuracy especially if the device used is extended by an ironsight. On the other hand, the user is forced to use his hands for holding the pointing device and so he is inhibited from using them for other kinds of interaction. Thus a unimodal gaze-based interaction approach can be worth using even for people without eye tracking experiences: it solves the just mentioned problem and provides acceptable results concerning accuracy and speed. However, it introduces new disadvantages concerning the usability: The user is limited in his movements because of the cable and the solution is less accepted concerning joy of use because of the uncomfortably eyetracking glasses. Using a more user friendly wireless eyetracker could improve the results especially concerning participants without experience of wearing an eyetracker.

7.1 Relationship to other studies

A set of related studies presented in section 2 shows the gaze-based method's advantage over established interaction methods [TJ00, DBMB07, CSD03]. What tells our study apart is the immersive CAVE environment used (see 3.2): Most other studies did not suffer from a great danger of displaced eye-tracking cameras because of extensive body movements. However, it is necessary to allow the user to move freely in CAVE applications for gaining a high level of immersion. Our study supports the opinion that eye-tracking could be used as an alternative to other interaction devices in VR [Pfe08], but on the other hand, one has to reflect the disadvantages concerning displaced cameras and limited freedom of the user's movements due to a cable.

7.2 Conclusion

To sum it up, cheap and established input methods like the pointing-based ones seem to offer an appropriate input for VR if the user is required to move freely and if he has his hands free for interaction: These methods are accurate, easy to explain, accepted also by novice users and easy to handle (no need for a special calibration). In contrast, gaze-based methods require expensive hardware, a personalized calibration and suffer from the risk of displaced cameras (and therefor a destroyed calibration). On the other hand, gaze-based methods are an important alternative if the user is handicapped or unable to use his hands for a certain action or if he is not required to use is head or his whole body. Especially in the second case, gaze-based approaches can provide a much higher immersion and performance than established ones [SG06].

7.3 Best Practices

After evaluating our studies, some best practices concerning input methods for VR were developed. The most important ones are:

1. Single eye blinks do not work as triggers in general; use both eyes.
2. Make the users comfortable in moving their heads, improvement of gear might be required.
3. Design interaction methods which are as related as possible to real life activities (e.g. using the Wii Remote with its pistol grip).

4. Visualize the user's interaction for reaching a high level of acceptance.

8 Future Work

Future research should reinvestigate the performance of the gaze-based interaction methods by using a wireless eyetracker which is more comfortable to wear and less likely to shift on the user's head. It would also be necessary to compare the developed methods with hand-based interaction as presented in [SZ94].

References

[Bol81] R.A. Bolt. Gaze-orchestrated dynamic windows. *Proceedings of the 8th annual conference on Computer graphics and interactive techniques*, pages 109–119, 1981.

[CH87] Ware C. and Mikaelian H. An Evaluation of an Eye Tracker as a Device for Computer Input. In *Proceedings of the SIGCHI/GI*, pages 183 – 188, New York, NY, 1987. ACM Press.

[CSD03] Nathan Cournia, John D. Smith, and Andrew T. Duchowski. Gaze- vs. hand-based pointing in virtual environments. In *CHI '03*, pages 772 – 773, New York, NY, 2003. ACM Press.

[DBMB07] M. Dorr, M. Böhme, T. Martinetz, and E. Barth. Gaze beats mouse: a case study. In *COGAIN*, 2007.

[IM06] Poika Isokoski and Benot Martin. Eye Tracker Input in First Person Shooter Games. In *The 2nd Conference on Communication by Gaze Interaction COGAIN 2006: Gazing into the Future*, 2006.

[OH08] Takehiko Ohno and Riad Ibrahim Hammoud. *Gaze-Based Interaction*, chapter 8, pages 181–193. Springer Series on Signals and Communication Technology. Springer, Berlin Heidelberg, 2008.

[Pfe08] Thies Pfeiffer. Towards Gaze Interaction in Immersive Virtual Reality: Evaluation of a Monocular Eye Tracking Set-Up. In *Virtuelle und Erweiterte Realität - Fünfter Workshop der GI-Fachgruppe VR/AR*, pages 81–92, Aachen, 2008. Shaker Verlag GmbH.

[PLW09] Thies Pfeiffer, Marc Erich Latoschik, and Ipke Wachsmuth. Evaluation of Binocular Eye Trackers and Algorithms for 3D Gaze Interaction in Virtual Reality Environments. *Journal of Virtual Reality and Broadcasting*, 5(16), jan 2009.

[RDS+10] Patrick Renner, Timo Dankert, Dorothe Schneider, Nikita Mattar, and Thies Pfeiffer. Navigating and selecting in the virtual supermarket: Review and update of classic interaction techniques. In *Virtuelle und Erweiterte Realität: 7. Workshop der GI-Fachgruppe VR/AR*, pages 71–82. Shaker Verlag GmbH, 2010.

[SG06] J. David Smith and T. C. Nicholas Graham. Use of eye movements for video game control. In *ACE '06: Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology*, page 20, New York, NY, USA, 2006. ACM.

[SZ94] D.J. Sturman and D. Zeltzer. A survey of glove-based input. *Computer Graphics and Applications, IEEE*, 14(1):30–39, 1994.

[TJ00] Vildan Tanriverdi and Robert J. K. Jacob. Interacting with eye movements in virtual environments. In *CHI 2000*, pages 265–272, New York, 2000. ACM Press.