Towards Continuous Gaze-Based Interaction in 3D Environments – Unobtrusive Calibration and Accuracy Monitoring

Patrick Renner, Nico Lüdike, Jens Wittrowski, Thies Pfeiffer

Artificial Intelligence Group
Faculty of Technology, Bielefeld University
Universitätsstraße 25, D-33615 Bielefeld
Tel.: +49 (0)521 / 106 - 6999
Fax: +49 (0)521 / 106 - 2962
E-Mail: {prenner, nluedike, jwittrow, tpfeiffe}@techfak.uni-bielefeld.de

Abstract: Gaze-based interaction in virtual reality promises to have interactional advantages, but the current state-of-art is still faced with usability issues. Two of them, the decrease in accuracy (drifts) under continuous usage, e.g. due to slippages of the gear, and the obtrusive standard calibration procedure are addressed in this work. We propose a new calibration procedure that blends smoothly in the virtual environment: an eye-catching moving object enters the field of view and while the user follows this object, the tracking is calibrated. In addition, we designed two heuristics to detect drifts automatically and thus trigger calibrations only when necessary. The applicability of both approaches was positively verified in a user study.

Keywords: Virtual Reality, Immersion, Eye tracking, Calibration, Accuracy Monitoring

1 Introduction

When measuring people's eye movements there is a frequent need to calibrate the eye-tracking system to compensate drifts [MR02]. Otherwise, interaction quality and visual quality might decrease. Drifts are especially a problem for gaze-based interactions of a moving user [Pfe08]. In such cases, periodical calibrations are necessary to maintain a sufficient accuracy. These calibrations are typically done either in static intervals, e.g. every five minutes or every ten items, or they are triggered by a human operator.

The standard calibration procedure presents a grid of targets which have to be fixated in a sequence. This procedure and the high frequency of calibrations disrupts the flow of interaction and decreases the sense of immersion. This problem becomes increasingly important as modern eye-tracking devices are hardly more disruptive than normal 3D glasses. We observed cases where people even forgot to hand the device back after a study.

We address these problems by aiming at a reduction of the negative effect of the calibration procedure on the ongoing interaction. Such an unobtrusive calibration procedure should provide the same accuracy as the standard procedure. In addition to that, we propose ways to maintain a required accuracy aiming at a reduction of calibration frequency. For this, we offer two heuristics, one in the 2D eye space and one in the 3D world space.

After the presentation of related work in the next section, an overview of the equipment and the setting is given. Then, in Section 5, the new approach for unobtrusive calibration and the accuracy monitoring heuristics are explained. Subsequently, Section 6 reports on the user study which verified our approaches. Section 7 discusses the methods and Section 8 presents opportunities for future work.

2 Related Work

Witmer and Singer define immersion as "a psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences" [WS98]. Immersion can increase task performance and help the user to concentrate on the task. Thus it is meaningful to enhance immersion when striving for efficient interaction in virtual environments. A loss in interaction accuracy and frequent obstrusive calibration procedures counteract this aim. The optimal solution for this problem would be using techniques that calibrate the eye tracker automatically or that do not require a calibration. This is particularly necessary in other context as well, e.g. for measuring eye movements of very young children or animals. These can not be instructed to look at specific points for calibration.

Zoccolan et al. [ZGC10] developed an eye tracker for rodents. In their approach a camera moves around the rodent's eye and captures its geometric properties. However, while the eye movements can be estimated from these parameters, it is not possible to specify the exact point of regard. Kohlbecher et al. [KBB+08] use stereo cameras. They assume that the gaze vector is parallel to the normal vector of the pupil, which can be estimated from the 3D representation of the eye. However, because of high costs, stereo cameras per eye are not installed in standard eye trackers. A comparison of several calibration techniques using a calibration grid is given by Ramanauskas et al. [RDD08]. They compared different grid sizes and underlying mathematical models. However, this work focuses on the achieved accuracy, but not on timing, obtrusiveness or immersion.

Stampe [Sta93] considered the problem of obtrusive calibration in 2D space. In their studies, each user has to solve a sequence of tasks and they assumed that calibration is not necessary after each task. Therefore the user was shown a grid (of letters in this case) that fitted the calibration grid. This way, they detected drifts by comparing gazes on the grid points with the calibrated points. In a derivative work, Stampe and Reingold used a dynamic drift correction [SR95].

A less restricted method for monitoring the accuracy in 2D space was developed by Hornof and Halverson [HH10]. In their approach, *required fixation locations*, i.e. points the user is instructed to look at or points expected to be looked at, are used to calculate inaccuracies in the calibration. The current distribution of fixations around these points is then used to

Angular accuracy	$0.25° - 1.0°$
Angular precision	$0.15°$
Temporal resolution	30/60Hz
Optical resolution	640x480/320x240 Pixel

(a) Technical data of the eye tracker

(b) Modified eye tracker

Figure 1: Eye-tracking equipment used for the study.

detect inaccuracies. A calibration procedure can be triggered if necessary.

The object-based monitoring heuristic for 3D space presented in Section 5.2 is inspired by the idea of Hornof and Halverson [HH10]. The concepts behind Stampe [Sta93] are included in the unobtrusive calibration method (Section 4).

3 Equipment and Setting

The techniques were developed in a 3-sided TRI-SPACE, a CAVE-like environment. The user is able to move around freely in its limited space, so interaction in it is semi-immersive. An eye tracker from Arrington Research was used. It was modified by adding polarized glasses for stereo view and a tracking target (Fig. 1(b)), which serves to adapt the viewpoint relatively to the position and orientation of the user's head [Pfe08]. The eye tracker's technical data can be found in Table 1(a).

For evaluation of the methods an immersive virtual scenario called "The Biosphere" (Fig. 2) was chosen. The Biosphere was developed by a students course in the winter term 2010 at Bielefeld University. It is inspired by the movie Avatar. It consists of huge flying rocks, which have different kinds of plants and animals with specific behaviours on them. The Biosphere resembles a natural landscape, aiming to achieve a high degree of immersion.

4 Unobtrusive Calibration

To calibrate an optical eye-tracking system, a mapping between the observed image of the eye and the point of regard in the field of view (here the virtual environment) has to be constructed. Usually, parameters for an underlying mathematical model are set (see [RDD08] for details). In state-of-the-art calibration procedures the user is shown simple geometries arranged on an uniform grid. The user then follows a sequence of grid point presentations for calibration. When the user focuses a point, he triggers a signal to the system, either manually by pressing a button or automatically by detecting the fixation, and the system proceeds with the next point. Once the model parameters are collected, the eye tracker then can compute the point of regard based on the current image of the eye relative to a 2D plane.

Usually, this procedure is done in front of the screen of the eye-tracking system. For this the user has to leave the virtual environment. In addition, the user is asked to keep his head still during calibration. We improved this situation in earlier work: The calibration grid is now shown in VR and head tracking allows the user to move his head freely [Pfe08].

However, the calibration grid either overlays the scene in which the current interaction takes place or the virtual environment is faded out and the grid is shown in front of a plain background. It is obvious that this procedure has a negative impact on the user's immersion.

To face this undesired effect, the idea of our unobtrusive calibration procedure is to blend in smoothly into the current VR scene, if possible. At the beginning of the interaction, the user is instructed to follow a specific calibration object with his gaze, whenever it appears in the scene. Aligned to our scenario, a fancy dragonfly was designed for testing and evaluation (Fig. 3). Whenever a calibration is necessary, the calibration object flies into view, pursuing a pattern along the original grid positions.

For a successful calibration, however, the user has to reliably follow the object. This can be facilitated by using an eye-catching appearance and lighting or sound effects to make the calibration object more prominent, if the user is not following the object. To ensure that the calibration succeeded, the position of the moving object and the current estimation of the point of regard can be monitored. This approach should be less obtrusive than the grid-based approach. It is, however, still possible that the user is made aware of the eye-tracking device. However, we expect that interaction overall will get more natural and more immersive.

5 Accuracy Monitoring

Variations of the gaze data can either result from an inaccuracy of the eye tracker or from a changed focus of attention by the user. The challenge is to differentiate between these possibilities.

To increase robustness, we propose two heuristics to detect the need for recalibration.

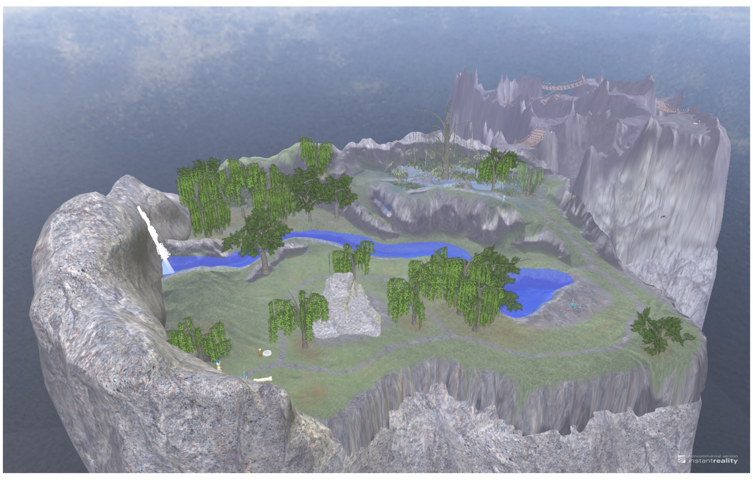

Figure 2: Our setting for the study is placed on a flying rock of our virtual Biosphere

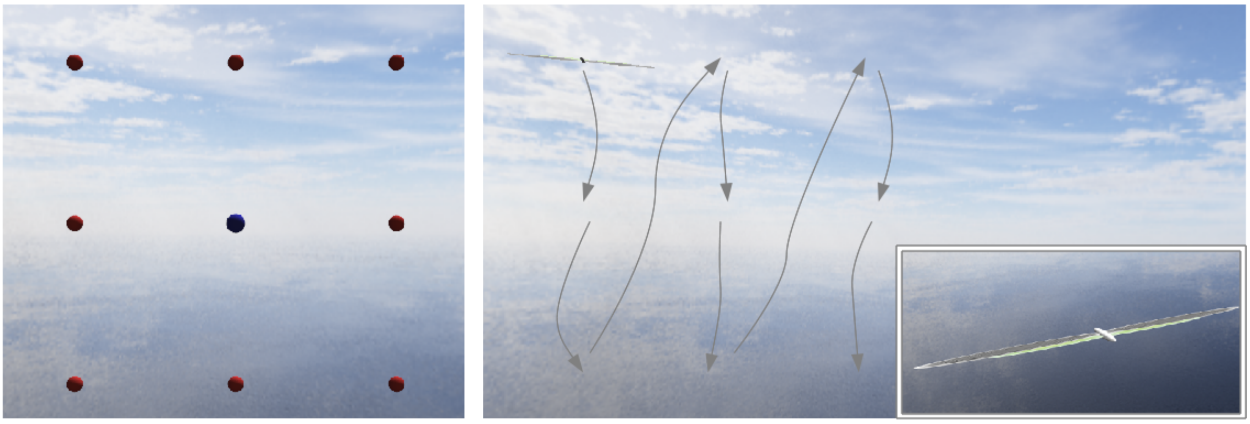

Figure 3: Classic calibration grid (left) and trajectory of the dragonfly (right) as moving calibration target. The dragonfly is depicted at the bottom right.

The heuristics are based on different assumptions, so if both of them detect an inaccuracy, the probability of a real need for calibration rises.

In normal gaze-based interaction, a threshold, e.g. defined by a desired accuracy, can be used to trigger calibration automatically. When the domain contains small features, high accuracy can be required, while the threshold can be relaxed if only the rough direction of gaze is relevant. The threshold could also be changed dynamically depending on visual context. Beyond this, dynamic calibrations are typically not wanted in user studies, as they could influence the performance metrics. However, knowledge about the gaze tracking accuracy during a trial might be useful, e.g. for quality assessment of the recorded data. Thus, the result of accuracy monitoring can be logged.

5.1 Eye-based heuristic

When people interact with virtual environments they look at all targets that strike their attention. These targets can be situated anywhere on the scene display, leading to a continuous movement of the eyes. However, the head is moved to compensate for eye movements beyond 30° from straight ahead. Hence we can expect that the angles between the eye gazes and the center eye gaze (which would mean that the user looks straight ahead) are relatively small and that the average eye gaze is almost constant, meaning that the distribution of the fixations is similarly balanced during the scenario of the application. This assumption can be used to estimate the calibration quality which is described in the next section.

Method: The method has two phases: training and monitoring. In training the typical fixation region of the user while interacting with the given scenario and task is learned. This region depends, e.g., on how often the user moves his head to compensate for eye movements. Therefore the eye gazes for the left and right eye of the person are measured during interaction inside the CAVE for three minutes directly after calibration, when accuracy can be considered high. After that, the average eye gazes of each eye are computed.

In the evaluation phase, the mean eye gaze is periodically computed for each eye and then compared to the reference values. As they are given in form of vectors we can determine the angles between them by using the dot product. The angles between the reference eye gaze and the mean eye gaze constitute the parameters to determine the calibration error. If the angles are beyond a threshold, there is a strong hint that the eye tracker needs to be calibrated.

Advantages & Restrictions: One advantage of the eye-based heuristic is that it is an intrinsic method. It can be easily integrated into any application that uses the eye tracker as a user interaction device, because it is decoupled from the application itself and uses no knowledge from the application's scenario. It even adapts itself to the scenario of the given application as it first learns a typical eye gaze distribution in the given scenario and uses that one as the reference to which it can compare later results. As the eye-based heuristic is based on the assumption that the average eye gaze is constant, errors will be caused when this assumption does not hold for the scenario of the application. For example, a combination of this heuristic and head-oriented steering is not useful.

5.2 Object-based heuristic

The object-based heuristic is guided by the idea that complex objects bear a restricted number of features that are target of visual attention. Consider for example virtual characters or animals. The user will probably mostly look at the face, as it is natural for humans. This quality of interesting objects can be used to monitor the eye tracker's accuracy. The object-based heuristic is an extrinsic heuristic. As information about the scene is used, this method can thus be considered to monitor the application-relevant accuracy of the eye tracking, which is effectively more important than the eye-based accuracy. It also includes the accuracy of the coupling of gaze and head tracking.

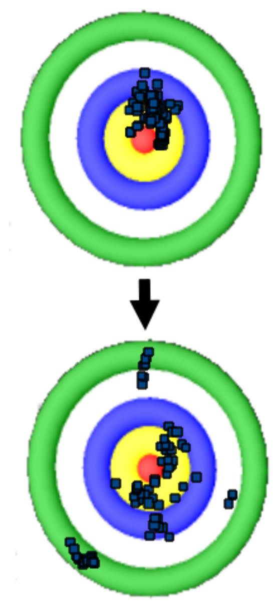

Figure 4: Fixations at target, beginning and end of interaction

Method: For each interestig object in the scene a "point of interest" (POI) is trained, i.e. the average of the points of the object the user looks at during interaction. This is done during the first minute after calibration, when accuracy can be considered high. If the user is explicitly instructed to look at a specific point of an object, the POI can also be set manually.

After that, data are recorded over periodic intervals (here: 80 seconds). After each interval, the mean of this data is compared to the trained POI (Fig. 4). The difference reflects the inaccuracy of the eye tracking. Angular measures are used to abstract away from the depth of the objects. When the difference exceeds a certain threshold, a possible inaccuracy is alerted.

For reasons of robustness, it is advisable to estimate POIs of more than one object in the scenario. All differences from the POIs can be centrally collected and compared. If a specific percentage of them exceed the threshold, an automatic calibration can be triggered.

Restrictions: In dynamic scenarios, in which the user is able to move around freely, this heuristic can only be used if the viewing angle from the user on the object stays in a certain domain. If the user, e.g., moves behind a virtual character, he will of course not look at the same POI as before. Actually, the former POI may not be visible anymore. In this case there are two possiblities:

1. if the eye tracker was calibrated only a short time ago: a POI for the new angle is learned

2. if accuracy can no longer be guaranteed, the object has to be ignored for accuracy monitoring

6 Evaluation

The evaluation was integrated in the Biosphere scenario (see above) in order to have an immersive surrounding. The users had to navigate to a special place in the Biosphere where the eye tracker was activated. At first it had to be calibrated. Half of the users calibrated using the standard calibration grid, the other half had to follow the dragonfly. After calibration targets where shown successively at exactly the positions of the grid points in order to assess the calibration's accuracy.

The users' task was to look at the middle of the targets (Fig. 5). After a short fixation on it, a target vanished and the next one was shown.

When the accuracy for every grid point was tested, the evaluation of the heuristics started. The procedure was similar to the former: Eight targets where shown successively, the users again had to look at the middle of each to make them vanish. The targets where shown in a loop. Each target was staffed with an object-based sensor, so gazes at them were recorded. In this case, a training of the POI was not necessary because the users were instructed to look at a specific point which was predefined as POI. Moreover the eye-based heuristic was activated at the start of this evaluation step.

Data were collected for about four minutes. Then the eye tracker was slightly dangled to cause a little inaccuracy (one side of the glasses was lifted a centimeter). After two minutes, the eye tracker was dangled a second time, but stronger (the cameras were touched). After another two minutes this evaluation part ended.

In the end, the user had to calibrate the eye tracker again, this time using the second calibration method, that had not been used in the beginning. The accuracy of that method was checked by showing targets as in the beginning. In addition, after the trial the user was interviewed about the two calibration methods.

Figure 5: Setting of the user study, the user is aiming at the bull's eye with his gaze

The scenario ran in real time and was not noticably decelerated by the monitoring heuristics.

6.1 General

The system was tested with 10 subjects, 9 male and 1 female (average age 23.2 years). The participants were mostly employees of the Artificial Intelligence group or members of other student projects held in this summer term. For this reasons, all of our subjects have already had experiences in virtual reality prior to the experiment and were at least roughly familiar with the field of eye tracking.

6.2 Data Evaluation

During the experiments raw data from the eye tracker were recorded as well as the accuracy after calibration and the outputs of the two monitoring heuristics. The results are as follows:

Accuracy of calibration methods: There is only a very little difference between the two calibration methods (Tab. 1). The classical grid method had an accuracy of 0.042 rad in median while the unobtruive method's accuracy was 0.036 rad in median with standard deviation of 0.015 rad and 0.016 rad. This indicates that the new method is at least as accurate as the classical one, but the t-test results in an error margin of 15%, so no assured evidence is given here.

Subject	Grid	Moving Object
1	0.040	0.040
2	0.076	0.034
3	0.042	0.070
4	0.037	0.051
5	0.043	0.041
6	0.061	0.030
7	0.047	0.023
8	0.041	0.022
9	0.019	0.038
10	0.039	0.018

Table 1: Accuracy of both calibration methods

Monitoring heuristics: After one minute of training, the eye-based heuristic calculated the amount of inaccuracy once per minute. Figure 6 (left) a) shows that a slight disturbance of the eye tracker barely caused

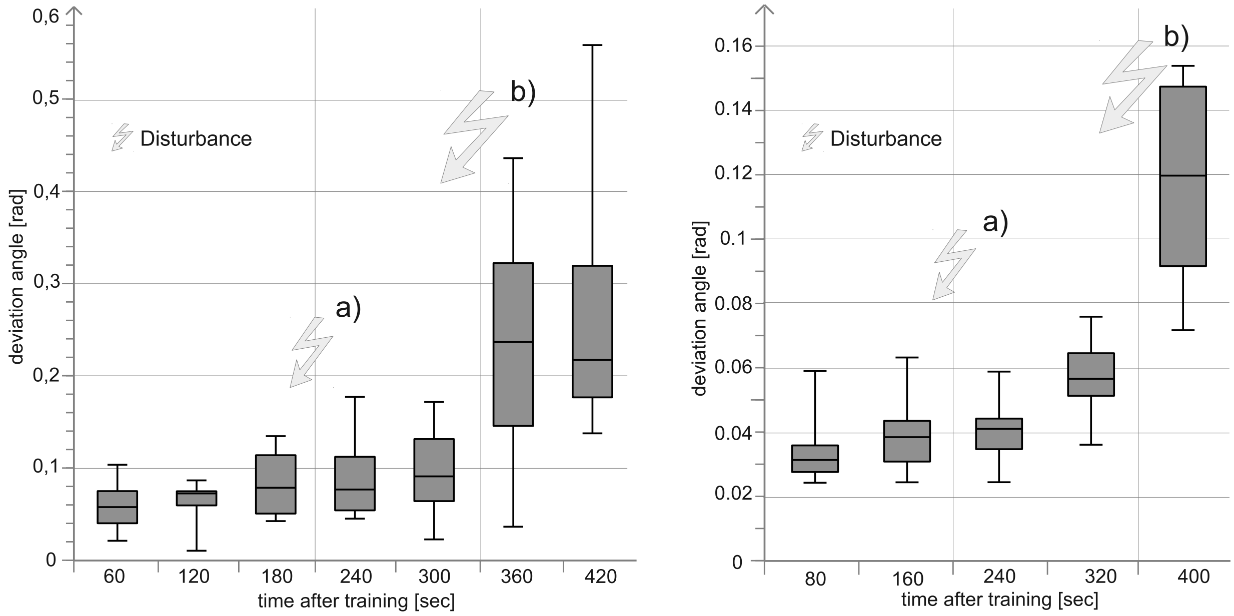

Figure 6: Left: Deviation from trained data of eye-based heuristic; Right: Deviation from trained data of object-based heuristic

a detection of inaccuracy. The object-based heuristic detects no distinct decrease in task performance, which suggests that this disturbance did not lead to a significant inaccuracy. However, the second disturbance (Fig. 6 b)) caused a distinct detection of the inaccuracy. The difference angle increased significantly from a mean of 0.095 rad before the disturbance to 0.230 rad after it (error margin with t-test $< 1\%$).

As mentioned above, eight targets for testing the object-based monitoring method were used. Figure 6 (right) shows the median of the difference angles between the expected point (the center of the targets) and the measured average point for each target. The first disturbance causes a slight increase of the difference angle in general (0.01 rad-0.02 rad). After the second disturbance all object-based monitors show a significant increase (Fig. 7) in their difference angles (mean of 0.057 rad to mean of 0.120 rad; error margin $< 2.5\%$).

6.3 User Interview

After the experiments the subjects were asked questions about how they experienced the calibration process, especially, how they judge the two presented calibration techniques.

The subjects had to give their answers on a scale from -2 (total disagreement) to +2 (total agreement). There were as well negative questions as positive ones ("Did you have fun during calibration" vs. "Did calibration disturb you"). These are the results (see Fig. 8):

Easiness: On average, unobtrusive calibration was judged to be easier (1.8 vs. 1.3, higher is better, not significant).

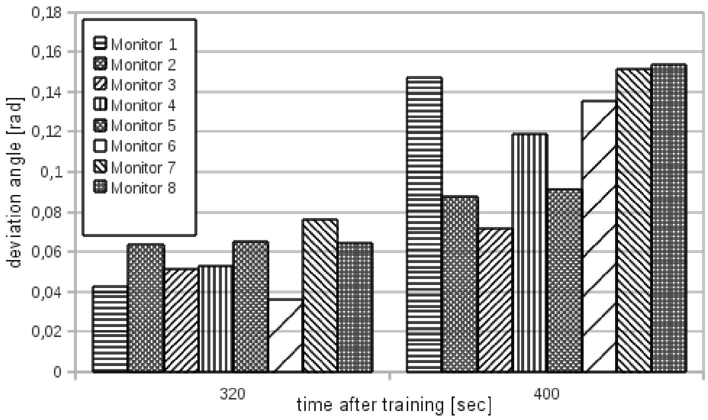

Figure 7: Outputs of the object-based monitoring for each target before (320 s) and after (400 s) harsh disturbance

Quickness: Regarding to efficient interaction, grid calibration was regarded quicker than moving object calibration (1.9 vs 1.3, higher is better, not significant).

Joy of use: All of the subjects agreed that unobtrusive calibration is fun (average of 1.6), while very few participants agreed to this regarding grid calibration (average of -0.7). A u-test assured this is highly significant (error margin $< 1\%$).

Disturbance: Users felt less disturbed by unobtrusive calibration in comparison to the classical calibration technique (-1.6 vs. -0.8, less is better, not significant).

Subjective accuracy: The users had the impression of good aiming performance using both calibration procedures. Unobtrusive calibration had slightly better result. (1.8 vs. 1.5, higher is better, not significant).

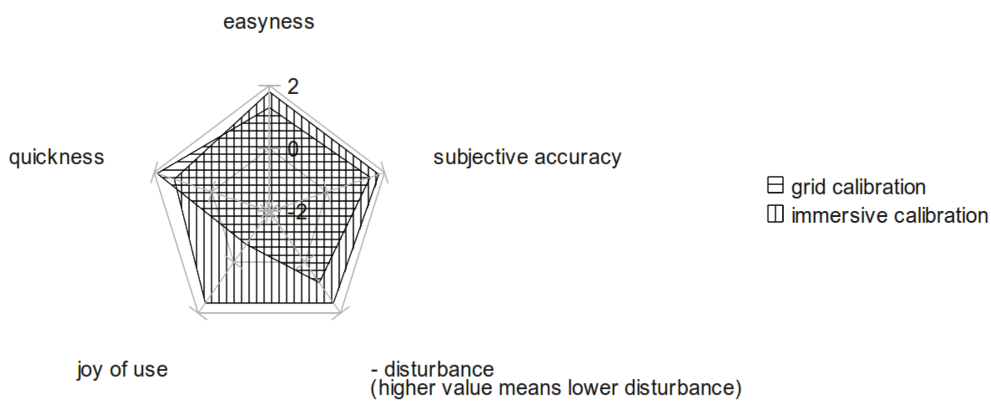

Figure 8: Overview of survey results

7 Conclusion

During gaze-based interaction in virtual reality immersion is often disturbed by frequent calibrations and the different appearance of the virtual world and the calibration grid.

Methods that attentuate this problem were developed: The unobtrusive calibration process improves user experience by using an object that is embedded in the scenario and blends in smoothly. Our results suggest that this method is easier to use than a standard calibration procedure. Users also have more fun using it, which is particularly important for long-term interactions.

Moreover it is possible to reduce the number of calibrations to a minimum. Two heuristics were developed for monitoring accuracy and reliably detecting inaccuracies. The eye-based heuristic concentrates directly on the orientation of the eyes, so it can be used independently of the application. The object-based heuristic considers interesting objects, on which the user's gaze will barely differ during interaction. Especially the combination of both heuristics ensures a good result. Calibration can thus be triggered automatically and only when it is really necessary. Very slight inaccuracies are barely detected, which can both be an advantage and a disadvantage, depending on which level of accuracy has to be reached and how often the user should calibrate the eye tracker.

In the current implementation, there are also some restrictions: Problems can occur when the user moves freely through a dynamic scenario. Potentially the eye-based heuristic will only work properly with long train- and test-intervals. A sudden change in the landscape can cause the heuristic to mistakenly detect an inaccuracy. The object-based heuristic can only be used as long as the angle to the object does not change significantly.

In general, the unobtrusive calibration method and the accuracy monitoring can positively influence the user experience and speed of an interaction with eye tracking. Moreover, they are able to help maintaining the user's motivation.

8 Future Work

The evaluation shows that the unobtrusive calibration method and the two heuristics for monitoring accuracy can improve immersion in eye tracking settings and detect inaccuracies of the tracking device. Nevertheless, the methods can be enhanced by some modifications.

Both heuristics currently run in constant time cycles. A faster and more correct monitoring could be realised by using a sliding-window technique, that constantly compares current data to the training data. Instead of training a POI, the object-based heuristic could train a gaze pattern and use RANSAC [FB81] to detect deviations. This should also increase the monitoring's correctness. The eye-based heuristic offers another possibility. By using Principal Component Analysis [Jol05] the dislocation axes could be found. This would also arise the possibility of automatically correcting the eye tracker's data unattended, without a calibration by the user.

References

[FB81] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24:381–395, June 1981.

[HH10] A. J. Hornof and T. Halverson. Cleaning up systematic error in eye tracking data by using required fixation locations. *Behavior Research Methods, Instruments, and Computers*, 34:592–604, 2010.

[Jol05] I. Jolliffe. *Principal Component Analysis*. J. Wiley and Sons Ltd, 2005.

[KBB+08] S. Kohlbecher, S. Bardinst, K. Bartl, E. Schneider, T. Poitschke, and M. Ablassmeier. Calibration-free eye tracking by reconstruction of the pupil ellipse in 3d space. In *Proceedings of the 2008 symposium on Eye tracking research*, ETRA '08, pages 135–138, New York, NY, USA, 2008. ACM.

[MR02] P. Majaranta and K. Räihä. Twenty years of eye typing: systems and design issues. In *Proceedings of the 2002 symposium on Eye tracking research & applications*, ETRA '02, pages 15–22, New York, NY, USA, 2002. ACM.

[Pfe08] T. 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. Shaker Verlag GmbH, 2008.

[RDD08] N. Ramanauskas, G. Daunys, and D. Dervinis. *Investigation of Calibration Techniques in Video Based Eye Tracking System*, pages 1208–1215. 2008.

[SR95] D. M. Stampe and E. M. Reingold. Selection by looking: A novel computer interface and its application to psychological research. In R. Groner and G. d'Ydewalle, editors, *Eye Movement Research - Mechanisms, Processes, and Applications*, volume 6, pages 467 – 478. North-Holland, 1995.

[Sta93] D. Stampe. Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behavior Research Methods*, 25:137–142, 1993.

[WS98] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3):225–240, 1998.

[ZGC10] D. Zoccolan, B. J. Graham, and D. D. Cox. A self-calibrating, camera-based eye tracker for the recording of rodent eye movements. *Frontiers of Neuroscience 2010*, 4, November 2010.