Time- and Space-efficient Eye Tracker Calibration

Heiko Drewes
LMU Munich
heiko.drewes@ifi.lmu.de

Ken Pfeuffer
Bundeswehr University Munich
ken.pfeuffer@unibw.de

Florian Alt
Bundeswehr University Munich
florian.alt@unibw.de

ABSTRACT

One of the obstacles to bring eye tracking technology to everyday human computer interactions is the time consuming calibration procedure. In this paper we investigate a novel calibration method based on smooth pursuit eye movement. The method uses linear regression to calculate the calibration mapping. The advantage is that users can perform the calibration quickly in a few seconds and only use a small calibration area to cover a large tracking area. We first describe the theoretical background on establishing a calibration mapping and discuss differences of calibration methods used. We then present a user study comparing the new regression-based method with a classical nine-point and with other pursuit-based calibrations. The results show the proposed method is fully functional, quick, and enables accurate tracking of a large area. The method has the potential to be integrated into current eye tracking systems to make them more usable in various use cases.

CCS CONCEPTS

• **Human-centered computing → Graphical user interfaces**;

KEYWORDS

Eye tracker calibration; Smooth pursuits; Linear regression.

ACM Reference Format:
Heiko Drewes, Ken Pfeuffer, and Florian Alt. 2019. Time- and Space-efficient Eye Tracker Calibration. In *2019 Symposium on Eye Tracking Research and Applications(ETRA '19), June 25–28, 2019, Denver , CO, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3314111.3319818

1 INTRODUCTION

Since the early eighties there is a vision of controlling computers with the eyes [Bolt 1981]. Interfaces based on gaze are promising but at the same time challenging [Jacob and Stellmach 2016; Jacob and Karn 2003]. One such challenge is the need of a time-consuming calibration. The calibration data reflects the geometry of the setup, typically eye tracker and screen position, and individual attributes from the user's eyes. On a personal system this has to be done only once as the geometry and personal attributes do not change. On a public system designed for more users and everyday interaction, however, calibration is necessary for any new user. This is particularly challenging [Khamis et al. 2016a] for short interaction times. Examples include an ATM where the user can securely enter the PIN with gaze [Cymek et al. 2014; Pfeuffer et al. 2013], public displays where users browse through the content [Zhang et al. 2013, 2014b], and generally any computing device, graphical user interface, or user study that utilises eye-tracking.

The design of a calibration procedure is characterized by a balance between *time*, i.e. the duration of the procedure, and *space*, i.e. where the calibration points are located. The standard procedure shows static targets and the user fixates each target for a few seconds. Usually 5 to 16 targets are presented in distinct locations. In sum, increasing time and space leads to better accuracy, but longer duration is reported as tedious and affects the system's usability [Flatla et al. 2011; Pfeuffer et al. 2013; Villanueva et al. 2004].

Researchers explored methods to address calibration issues with more dynamic methods that exploit smooth pursuit eye movements [Blignaut 2017; Celebi et al. 2014; Khamis et al. 2016b; Pfeuffer et al. 2013]. These methods correlate eye and stimulus movements to infer whether the user attends to the target. Whereas the standard calibration procedure needs guidance of the user through instructions, pursuit calibration can happen unconsciously without assistance. Furthermore, a calibration based on smooth pursuits is generally quicker than a standard calibration [Pfeuffer et al. 2013].

In this paper, we investigate a regression-based method to extract calibration parameters from smooth pursuit movements and present a study comparing the method with the standard point-based and other pursuit-based calibration methods. In comparison to other pursuit-based calibration methods, this method uses regression to analyze the scaling and translation parameters of the calibration mapping. The method has been introduced in [Drewes et al. 2018]. In this work we focus on the method for the purpose of gaze calibration from a theoretical and empirical perspective.

The method comes with two main benefits that we investigated in the study. First, the method is *time-efficient*, as it allows for calibrating within a few seconds. This is useful for public and mobile systems, where users interact frequently but in brief sessions. Second, the method is *space-efficient* as only a small area is required to calibrate. Yet, it enables gaze-tracking of a large area around it. This can be useful, for instance, when calibrating on a small screen of a mobile device but using eye-tracking outside the display and/or other screens around the user. Accurate calibration for the surroundings of the display is valuable for a new generation of mobile devices which use inside-out tracking for an awareness of the space around it. This means that the gaze coordinates outside the display can be mapped to objects in the surroundings.

The result of the study is that quick pursuit calibration is possible in 2 seconds in a small area, but with careful consideration of a trade-off between calibration speed and accuracy. In addition, the study helps to understand the strength and limitations of the regression, offset, and homography-based calibration mappings.

ETRA '19, June 25–28, 2019, Denver , CO, USA

ACM ISBN 978-1-4503-6709-7/19/06…$15.00
https://doi.org/10.1145/3314111.3319818

2 RELATED WORK

Researchers identified the calibration task as a significant problem for a greater adoption of eye gaze. Morimoto and Mimica's survey identifies calibration as a task of poor usability [Morimoto and R.M. Mimica 2005]. Schnipke and Todd's gaze and calibration tests report low accuracy of eye-tracking data [Schnipke and Todd 2000], and other researchers have emphasized the low usability of the 'tedious' calibration procedure that users need to conduct [Flatla et al. 2011; Pfeuffer et al. 2013; Villanueva et al. 2004].

It is possible to avoid the calibration by using calibration-free gaze interaction techniques. An early approach is are gaze gestures [Drewes et al. 2007; Drewes and Schmidt 2007]. Other approaches focus on directional scrolling [Zhang et al. 2014a,b] or use visual saliency to predict gaze [Sugano and Bulling 2015]. A recent approach utilizes smooth pursuit movements of the eyes [Vidal et al. 2013a]. If the gaze follows the pursuit target, the coordinates of the target and the coordinates reported from the eye tracker correlate. This method works without calibration, but requires targets in motion. To enable the default eye tracking without targets in motion, the calibration requirement remains.

To tackle the calibration challenge, researchers aimed to improve the default point-based calibration. Geometric models were investigated that allow the number of points required to calibrate to be reduced and thus a shorter procedure duration to be achieved. For example, [Villanueva et al. 2004] and [Ohno and Mukawa 2004] developed calibration procedures that only require 2 points to estimate gaze. Other variants involve 1-point calibration [Guestrin and Eizenman 2008]. Nonetheless, to achieve a high accuracy in practice, gaze calibrations in research and industry use additional points [Tob 2018; Villanueva and Cabeza 2008], that at the same time increase the duration of the procedure.

Alternatively, a moving target calibration based on smooth pursuit was investigated. By using the Pursuits approach [Vidal et al. 2013a] a system can infer gaze towards a moving target in uncalibrated environments. PursuitCalibration [Pfeuffer et al. 2013] extends the idea to calibration. It uses moving stimuli and collects sample data only when users follow the target. This allows to infer attention even during the procedure and improves its stability.

The method has been extended to different environments and calibration variants. Celebi et al. extended the method with a spiral movement pattern and the Bayesian-Gaussian regression technique for lag correction and outlier rejection [Celebi et al. 2014], finding that it improves accuracy and stability of the gaze samples. The CalibMe [Santini et al. 2017] method uses a similar approach to explore gaze calibration in pervasive environments using fiducial markers on screens to using spiral and star movement patterns for wearable devices. Their evaluation shows a higher accuracy than a 9 point calibration. Researchers also have used the method to provide continuous calibration during the user interaction sessions [Gomez and Gellersen 2018; Murauer et al. 2018; Ramirez-Gomez and Gellersen 2017; Tripathi and Guenter 2017].

Of particular interest are three pursuit based calibrations. First, Pfeuffer et al. have used a homography to establish the mapping between gaze and screen [Pfeuffer et al. 2013]. Khamis et al. used a simple offset correction method to establish the mapping as it reached sufficient accuracy for the tested scenario [Khamis et al. 2016b]. Finally, Drewes et al., in their investigation of smooth pursuit detection, pointed to the potential of using a regression-based method for gaze calibration [Drewes et al. 2018], however, used it only for presenting images of gaze trails. Our research focuses on these three methods that we will elaborate on in detail next.

3 THEORETICAL BACKGROUND

3.1 Calibration

The calibration of a stationary eye tracker means to find a mapping from coordinates reported by the eye tracker to coordinates on a screen. For a mobile eye tracker, the coordinates are mapped to coordinates on the image from the world camera.

The standard approach to calibrate an eye tracker is a linear transformation in homogeneous coordinates, called a homography. If calibrating to a flat screen, a linear transformation is not exact. An exact mapping needs trigonometric functions and is non-linear. For small displays the involved angles are small and as sin(x) ≈ x and cos(x) ≈ 1 for small angles, a linear approximation is legitimate.

Current eye tracking technologies need a calibration to the individual using the eye tracker. The position of a glint from an infrared LED on the cornea depends on the size of the eye ball, which is different from person to person. Additionally, the exact position of the fovea differs from person to person.

3.2 Correction of Calibration Errors

The manufacturer of an eye tracker normally provides a calibration procedure which establishes the internal mapping from the eye tracker camera to the screen. After calibration, the calibrations errors should be minimal and there should be no need for correction. However, if another person wants to use the eye tracker, a new calibration is necessary. This is an obstacle for instant use as public interface. Therefore, many researchers think about implicit and quick calibration procedures. Such implicit calibration does not necessarily do a mapping from the eye tracker's camera coordinates to the screen but can be a mapping on top of the existing calibration.

A full calibration compensates for geometric aspects, such as the position and orientation of the eye tracker against the screen, and for individual aspects of the user, such as the size of the eyeball. Within a scenario of instant use by different people the geometric aspects do not change as screen and eye tracker are still in the same position. The interesting question for this research is: what are the implications for the calibration process if we only swap the user but leave the geometrical setup unchanged?

The research presented here assumes an eye tracker calibrated to a person but used by another person. Consequently, we need a mapping from reported gaze coordinates to true gaze coordinates. [Pfeuffer et al. 2013] assume that this can be achieved by a linear transformation in homogeneous coordinates or so-called homography. The received coordinates r are multiplied with matrix H and the results are the true coordinates p (see Equation 1).

$$\begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \begin{pmatrix} r_x \\ r_y \\ 1 \end{pmatrix} \tag{1}$$

It needs at least four pairs of received and true screen coordinates to estimate the parameters of matrix H, as the matrix has eight degrees of freedom. The coordinate pairs are gathered in a calibration procedure and, as measurements always have some inaccuracies, it is better to have more than four coordinate pairs. This means there is an overestimated linear equation system to be solved. [Pfeuffer et al. 2013] used the RANSAC algorithm as it performs well on outliers and the estimation of homography parameters is a standard task in computer vision. The homography can be decomposed in scaling in x- and y-direction, rotation, shearing, translation in x- and y-direction and two trapezoidal transformations (Figure 1).

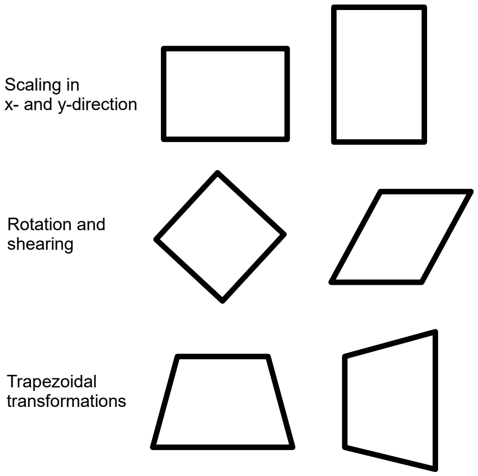

Figure 1: Beside translations in x- and y-direction a homography can be decomposed in scaling in x- and y-direction, rotation, shearing, and two trapezoidal transformations

One question of this research is which parts of the homography contribute how much to the correction of the calibration error and whether a homography is sufficient. Khamis et al. did calibration with pursuits and only corrected the offset which means they only used the translation part [Khamis et al. 2016b], which worked quite well. This means that the translation part contributes substantially to the calibration correction.

If an eye tracker was calibrated to a reference person and another person is using it, then there seems to be no reason why shearing and rotation should be necessary to correct calibration errors. Shearing occurs if x- and y-detection are not perpendicular. However, the grid on the photo sensor in the eye tracking camera is very precise and does not differ from 90°. Rotation is a question of the angle between the screen and the eye tracking device and this does not change with a different person in front of the system. The non-linear part of the homography is responsible for a perspective view which means it can transform a square to a trapezium. Such a transformation depends on the position of the eyes and, again, the situation does not change with a different person in front of the system. Consequently, a recalibration on another person only requires scaling and translation.

3.3 A New Calibration Approach Based on Linear Regression

The idea of this work is to use linear regression to estimate two scaling and two translation parameters. With four parameters this approach is in between the suggestion of [Khamis et al. 2016b], who used only two translation parameters, and [Pfeuffer et al. 2013], who used a homography with eight parameters. Linear regression uses data pairs, plots them in a plane and estimates the best fitting line for the plotted data. The data pairs are one coordinate of the gaze signal and the corresponding coordinate of the target. If there is no calibration error, target and gaze are in the same position and, therefore, have the same values for their coordinates. This means the fitting line is the bisectrix of ordinate and abscissa (Figure 2).

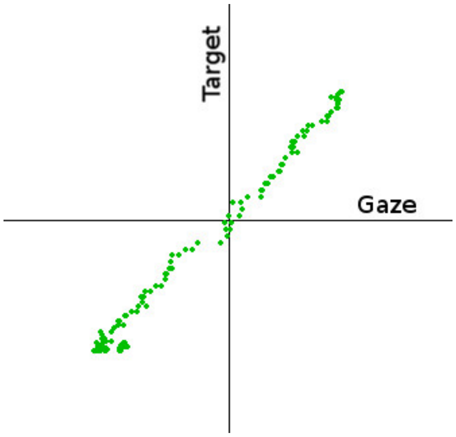

Figure 2: Regression analysis of a calibrated eye tracker signal. As gaze and target coordinates have the same values the slope of the fitting line is 1 and the intercept is 0.

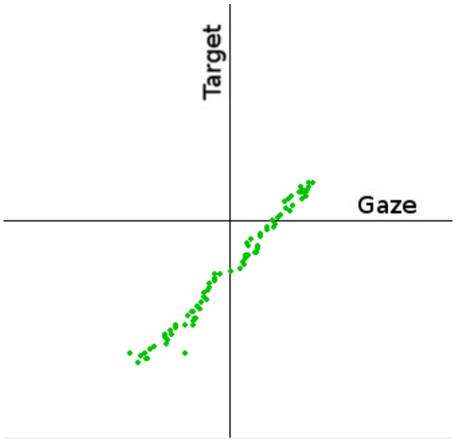

Figure 3: Regression analysis of uncalibrated gaze data. The intercept represents the offset and the slope the scaling.

In case of a calibration error the fitting line's slope is the scaling factor s and the intercept is the translation t needed for correction (Figure 3). Let t be the target coordinate and g_r the reported gaze coordinate. We assume that the eye follows the target exactly and therefor the true gaze coordinate g_t and the target coordinate are the same. As the true gaze coordinate can be calculated from slope s and intercept o we get the following relation:

$$t = sg_r + o = g_t \tag{2}$$

The linear regression analysis provide the parameters s and o. If using only translation and scaling as explained above the transformation matrix looks like this:

$$\begin{pmatrix} s_x & 0 & t_x \\ 0 & s_y & t_y \\ 0 & 0 & 1 \end{pmatrix} \tag{3}$$

If only the offset is considered, the matrix is even simpler:

$$\begin{pmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{pmatrix} \tag{4}$$

3.4 Calibration Quality

Figure 4 illustrates the meaning of the terms precision and accuracy as given in Tobii's white paper[1].

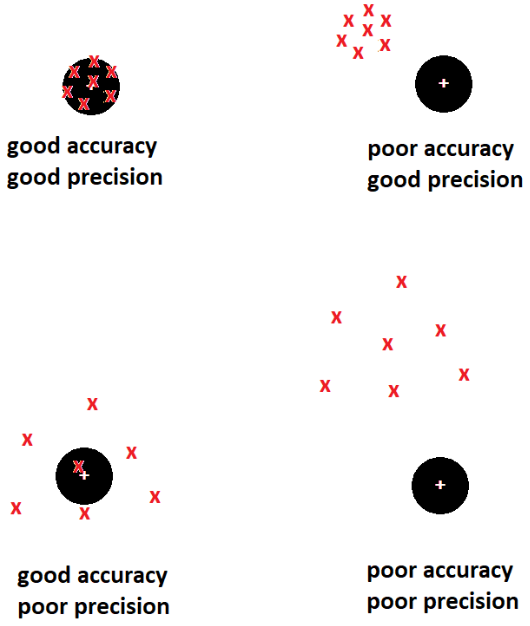

Figure 4: Illustration of accuracy and precision. The black circles are targets. The red crosses are gaze positions.

In terms of mathematics, accuracy is the distance from the gaze positions' mean or center of mass to the target center. If t is the target position and g_i are the gaze positions then the accuracy A for a static target is:

$$A_{statictarget} = \frac{1}{n}\sum_{i=1}^{n} g_i - t \tag{5}$$

There are several definitions of precision. One possible definition is the standard deviation of g_i. Another possibility is the average distance of the gaze positions to the gaze positions' mean.

It is obvious that the correction of calibration errors improves accuracy but does not change precision. To be precise, the precision is influenced by the scaling, but the scaling factors here are close to 1 and do not basically change the situation.

The situation for calibration with moving targets, i.e. with smooth pursuits, requires changes in the definition. As the target is moving there are corresponding target positions t_i for the g_i and the accuracy A moving target is:

$$A_{movingtarget} = \frac{1}{n}\sum_{i=1}^{n}(g_i - t_i) \tag{6}$$

A definition for the precision can be done in an analog way. Although the approach is similar there are subtle differences. Figure 5 shows two cases of target (black dots) and gaze (red crosses) positions for a circular pursuit task. Both cases show a perfect accuracy and both have the same precision.

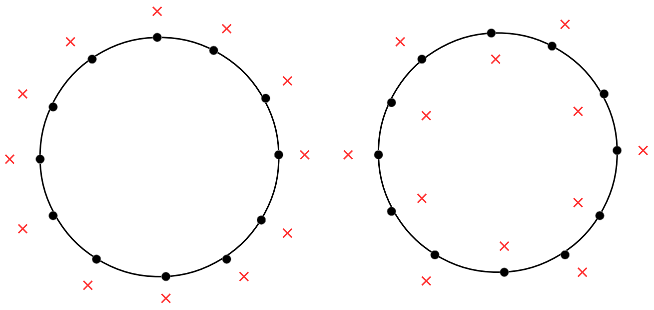

Figure 5: Illustration of accuracy and precision for smooth pursuit calibration. The black dots are the target position and the red crosses are the gaze positions.

However, for the left case it is possible to achieve a better calibration by scaling while this is not possible for the right case. This means we could need other definitions for accuracy and precision for pursuit tasks. This is not trivial as there are not only spatial but also time aspects. What should we state as accuracy and precision if the gaze is exactly on the target position but always one step behind? We leave this question for future work, as our main interest is on the calibration methods that we study next. There, we use the default definition given in Equation. 6.

---

[1] https://www.tobiipro.com/siteassets/tobii-pro/accuracy-and-precision-tests/tobii-accuracy-and-precisiontest-method-version-2-1-1.pdf

4 USER STUDY

We compare the three pursuit-based calibration methods: the offset, linear regression, and homography method. A nine-point standard calibration was used as the baseline as this is widely used. We decided to use circular trajectories as they have shown as promising for accurate pursuit detection [Celebi et al. 2014; Esteves et al. 2015].

4.1 Pilot Study

Our intention was to use a quick moving target on a small trajectory to achieve a fast calibration procedure. We implemented a first version of the system and asked 3 colleagues to complete a test run. The offset method and the linear regression method worked well, but the homography method did not provide reasonable results. Varying the RANSAC parameters in the homography function provided by OpenCV and also changing the calculation to least-median or least-square did not solve this problem. We found the homography method gives only good results for the area enclosed by the target trajectory. Or in other words, the homography method only works if the trajectory of the pursuit target is close to the edges of the display. This lead us to investigate multiple pursuit sizes.

4.2 User Study Design

The study uses a within-subject design with repeated measures.

The *first goal* of the user study was to show that a quick calibration with smooth pursuit eye movements is possible. For that, we use two different speed conditions to examine the influence of speed to the calibration accuracy. [Drewes et al. 2018] found that lower speeds achieve a better accuracy, while higher speeds enable a faster calibration. To cover both, we implemented one rotation in 2 seconds as an example for a ’very quick calibration’, and the other in 4 seconds where the target moves slower on the circle.

The results from the pilot study added a *second goal*, which was to understand when and why the homography method fails. Consequently, we used two different radii for the pursuit trajectory: a small trajectory for a fast calibration on a small area and a large trajectory which is close to the display edges. The radii were 200 pixels and 450 pixels or 4° and 9° respectively (Figure 6). The 200 pixels were a compromise between having a small area but still getting sufficient accuracy. The 450 pixels radius was chosen as it covers almost the whole screen area.

Overall, this leads to the following design:

3 *Calibration methods:* offset, regression, homography
2 *Sizes:* 200 px (4°), 450 px (9°)
2 *Speeds:* 2s, 4s
= 12 Conditions.

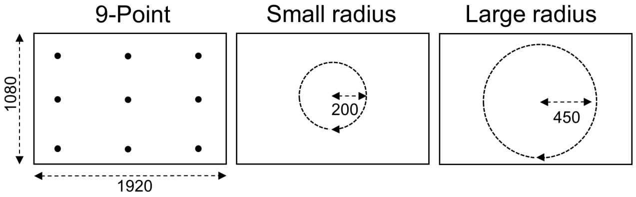

Figure 6: Design of the study calibration tasks.

4.3 Accuracy, Precision, and Outliers

Dependent variables are the accuracy and precision of the calibrations. The calculation of the homography uses the RANSAC algorithm for outlier removal. Accuracy and precision values are more exact if calculated without outliers. However, for comparison of these values with the pursuit calibration methods there should be outlier removal for the pursuit calibration, too. For the regression method it is possible to apply a RANSAC algorithm for outlier removal, but it is not clear whether the same threshold value or the same percentage of outliers is better for comparability. Also for the offset method it is possible to remove outliers but again the question arises which method should be applied.

For the three pursuit calibrations (homography, offset, regression), we calculate precision and accuracy from the data without outlier removal. This makes sure that we can compare the methods against each other and that different values would be specific to the method and are not caused by outlier removal.

4.4 Implementation

Our system consists of a laptop with built-in eye tracker (Tobii IS4 Base AC) which delivers gaze coordinates at 60 Hz. The display has a resolution of 1920 × 1080 px on 38.4 cm × 21.7 cm, which results in 0.2 mm for one pixel or 50 px per centimeter. The average distance of the participants’ eyes to the display was 50 cm ± 5 cm, corresponding to a visual angle of 0.02° per pixel (50 px per degree).

We wrote a program which offers a nine-point calibration and a pursuit target procedure. The nine-point calibration uses targets with a size of 50 px or 1° visual angle which means the target center of the eight outer targets have 25 px (0.5°) distance to the display edge. Each target is displayed for two seconds, resulting in 18 seconds for the nine-point calibration. The gaze data of the first 500 ms are excluded, to avoid the initial moving time to the target. This aligns with [Krassanakis et al. 2016] finding that initial target search and fixation on a simple UI takes 362±119 ms.

The pursuit calibration also added 500 ms to give the user time to catch up the target. The pursuit target is displayed for 2.5 seconds or 4.5 seconds but again the first 500 ms are not used for calculations.

The program calculates a homography mapping for the 9-point-calibration data, and the accuracy and precision for the homography corrected gaze data. For the pursuit tasks the program applies three correction methods – offset, regression, homography – to the gaze data from the raw gaze samples collected in the 9-point-calibration, and calculates accuracy and precision. Notably, having more than 9 points may improve accuracy overall, but have little effect on relative differences between conditions which is our main interest.

4.5 Participants and Procedure

We invited 16 participants to our user study, nine male and seven female, aged 23 to 54 years. Eight of them wore corrective glasses during the study.

After filling a form with their demographic data, all users started with the 9-point calibration. To make sure that the participant looked at the targets, the program calculates the average distance of the gaze points to the gaze points’ mean. If this average distance was beyond 60 pixels (1.2°) the participant was asked to repeat the 9-point calibration. This happened to one third of the participants.

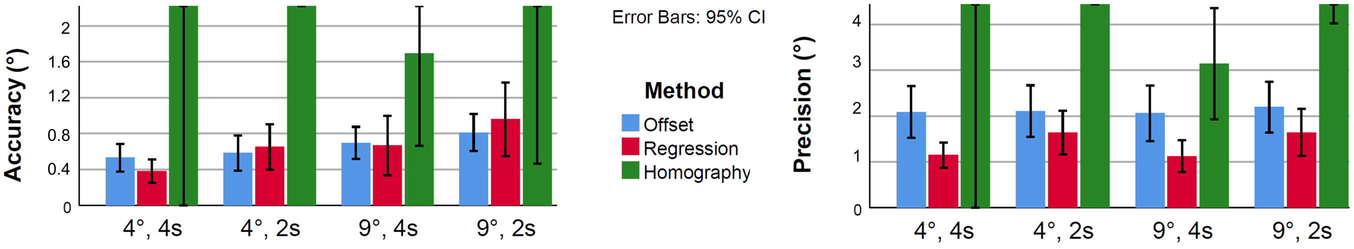

Figure 7: Study results on accuracy and precision for the calibration methods.

Then, users followed the pursuit targets in the four conditions. The order of conditions was randomized for each participant. Again, the gaze input was checked for plausibility by calculating the correlation. If the correlation in either x- or y-direction was below 0.9 the input was declared invalid and the participant had to repeat the pursuit task. However, this happened to none of the participants.

5 RESULTS

Table 1 shows the accuracy and precision of the 9-point calibration. We calculated these values considering also outliers as we need these values for comparison with the pursuit calibration. Calculating the accuracy of a data set from which the calibration parameters were calculated should have resulted in a perfect accuracy of zero. However, the outlier removal of the RANSAC algorithm has contributed to a slight deviation.

Notably, considering the 9-point calibration accuracy results with the same calibration samples, the results are less meaningful and should not be compared with the pursuit-based calibration. However, the precision data should be valid as we can assume the scattering of gaze data will not deviate much by the calibration. Therefore, the calculated precision has some meaning.

We calculated the precision as the average distance of gaze points to the target positions. Figure 4 suggests that accuracy and precision are independent from each other, but only if the accuracy becomes zero by calibration. In this case, however, the accuracy is not equal zero and this means that the value for the accuracy adds to the precision value. Figure 7 shows accuracy and precision that are measured by using gaze data collected from the 9-point calibration.

Overall, we found the offset and regression methods reached a better accuracy and precision than the homography-based approach. The homography is, under most conditions, substantially less accurate with values ranging from 84.8 px (1.7°) to 375.3 px (7.5°). The offset and regression show similar stable results ranging from 19 px (0.38°) to 48 px (0.96°). This shows that these methods are more flexible than the homography-based methods across conditions. Only in the condition of larger size (450 px/9°) and longer procedure (4 s), the homography resulted in an acceptable accuracy of 84.8 px (1.7°).

To analyse the differences in more detail, we consider the homography as an outlier and focus our statistical analysis on the offset and regression conditions. We conducted an ANOVA test with Bonferroni and Greenhouse Geiser corrections on the accuracy and precision, respectively, with the following design: 2 sizes (200 px, 450 px) × 2 times (4 s, 2 s) × 2 methods (offset, regression).

9-point	accuracy (std. dev.)	precision (std. dev.)
	0.1° (0.11)	0.43° (0.19)

Table 1: Accuracy and precision for the 9-point calibration (including outliers).

Size, Duration	offset	regression	homography
4°, 4s	0.53 (0.29)	0.38 (0.25)	5.25 (11.6)
4°, 2s	0.58 (0.37)	0.65 (0.48)	6.39 (4.98)
9°, 4s	0.70 (0.34)	0.67 (0.62)	1.70 (1.94)
9°, 2s	0.81 (0.39)	0.96 (0.77)	7.51 (13.2)

Table 2: Accuracy results (in visual angle) across conditions and methods. The standard deviation is given in brackets.

Size, Duration	offset	regression	homography
4°, 4s	2.09 (1.06)	1.14 (0.51)	7.28 (15.25)
4°, 2s	2.11 (1.05)	1.64 (0.89)	11.092 (10.12)
9°, 4s	2.062 (1.13)	1.12 (0.65)	3.14 (2.28)
9°, 2s	2.198 (1.04)	1.64 (0.96)	12.52 (15.94)

Table 3: Precision results across conditions and methods. The standard deviation is given in brackets.

5.1 Accuracy

The values for the accuracy are provided in Table 2. A graphical visualization is shown on the left side of Figure 7. The analysis revealed that both main effects *size* (F^1_{15}=7.3, p=.016) and *speed* (F^1_{15}=9.76, p=.007) reveal significant differences. This shows an expected result that with a larger path and longer time, the accuracy improves. Factor calibration *method* has not been found as significant (F^1_{15}=01, p=.9), which aligns with the smaller differences between offset and regression method (Figure 7a). The interaction effect between *speed* × *method* has been reported as significant (F^1_{15}=F, p=.033). Pairwise comparisons show that for the regression method, the accuracy was significantly better in the slow speed condition (p=.004), whereas no differences were found for the offset method (p=.17).

5.2 Precision

The values for the accuracy are provided in Table 3. A graphical visualization is on the right side of Figure 7. Factor *speed* has been found as significant (F^1_{15}=11.75, p=.004), i.e. as expected slower calibration speed leads to more precise gaze. Factor *size* has not been reported as significant (F^1_{15}=.013, p=.91), showing that it has less impact on precision. A significant effect was found for factor *method* (F^1_{15}=13.72, p=.002), showing that the regression method resulted in a significantly better precision than the offset method. Lastly, the interaction effect between *speed* × *method* has been reported significant (F^1_{15}=6.38, p=.023). Pairwise comparisons show that within both the offset (p=.044) and regression method (p=.008), the precision is better at the slower speed.

6 DISCUSSION

We explored calibration methods from a theoretical and empirical perspective. Here we summarize and discuss our main findings:

The regression and offset methods allow for time- and space-efficient calibration. The overall goal of the study was to show that it is feasible to calibrate in a short time in a small area. Our study shows that an accurate calibration has been established after 2 seconds of time using the regression or offset method. Both also allow to calibrate in a small area of calibration (radius: 4° visual angle).

The regression line and offset calibration give better calibration accuracy. Accuracy values below 1° are good enough for practical use. However, also for these methods the standard deviation is high which means for practical use an accuracy of the mean plus the standard deviation should be assumed. From the data given in Figure 7 the accuracy is 1° to 2°. Across the tested conditions, we find the slow small pursuit condition works best. A lower pursuit speed improves both accuracy and precision.

The homography calibration does not lead to a usable calibration quality. An accuracy in the range of 5° to 7° is not good for practical use. Additionally, the standard deviation is very high which means that the method is not very reliable. From the theory we expected that the homography calibration works better for the pursuit task with the big trajectory. This is true when the target moved slowly, but when faster the accuracy degraded substantially.

Regression is slightly more precise than offset and substantially more precise than homography. The regression and offset methods work clearly better than homography. The precision values are two to four times bigger than the 21.4 pixels (0.43°) precision from the 9-point calibration. The reason for this is that the lower accuracy has an influence on the average precision distance. It seems that the regression method has better precision than the offset method. This may be a consequence from the inclusion of scaling factors.

7 CONCLUSION

For the pursuit calibration methods we learned in this study that the newly introduced regression-based calibration method performs a bit better than the offset method, especially in the small slow pursuit condition. The homography method, however, only works well in the area of the pursuit trajectory.

The reason why the homography does not work well in the outer areas lies in the nature of the homography that involves many parameters of which not all are useful. When collecting gaze data for calibration there will be always some noise in the data created by tremor and micro-saccades of the eye and from technical limitations of the eye tracker. This noise propagates to the parameter values. Especially the noise induced error in the rotation parameter causes a calibration error for coordinates far away from the pursuit trajectory. The homography is more useful to compensate the geometry of eye tracker and screen. This leads to two different calibration scenarios. One scenario is a factory setup calibration which compensates the geometry. This situation needs a homography-based calibration with targets in the edges of the tracking area. The other scenario is a calibration for another person, where the geometry does not change but the eyes of the user. To calibrate here with minimal time and space needs, the regression or offset method is well suited.

The study suggests that the 9-point calibration produces a better result for accuracy and precision than the pursuit calibration methods. The price for the better accuracy is a longer time for the calibration process, in this study 18 seconds against 4 or 2 seconds, respectively. Additionally, the 9-point calibration requires the cooperation of the user and is perceived as a calibration process. A pursuit calibration, however, can happen without the user being aware of it. In these scenarios, the calibration target may not be tied to a circular stimulus and can be any path as well as any type of object to follow with the eyes (e.g., game objects [Vidal et al. 2013b] or text [Khamis et al. 2016b; Pfeuffer et al. 2013]), which needs to be evaluated in future work. A further future direction is to also consider different calibration aspects beyond the methods we investigated, such as polynomial transformation mappings of the gaze to the screen, and continuous calibration [Gomez and Gellersen 2018; Murauer et al. 2018; Tripathi and Guenter 2017].

Our findings are interesting for various applications, for example, for mobile devices such as smart watches or smart phones. Such devices could be calibrated at manufacturing time for a tracking area much larger than the size of the device. The buyer of this device could do a personal regression-based calibration with targets on this small device. Then, this device could track the gaze accurately, even for positions outside of the device’s display. For a public display which can be controlled by gaze, this means that the interaction elements have to be larger than on a personal display, which is calibrated on the person with a 9-point calibration. In the case of a public ATM, where the implicit calibration is hidden in an advertisement, larger interaction elements to enter a PIN by gaze are not a real obstacle. A time-consuming calibration, however, is a real obstacle on a public display, that can be alleviated using the methods we investigated in this paper.

REFERENCES

2018. Tobii SDK Calibration. http://developer.tobiipro.com/commonconcepts/calibration.html. Accessed: 2018-12-14.

Pieter Blignaut. 2017. Using smooth pursuit calibration for difficult-to-calibrate participants. *Journal of Eye Movement Research* 10, 4 (2017). https://bop.unibe.ch/JEMR/article/view/3676

Richard A. Bolt. 1981. Gaze-orchestrated Dynamic Windows. *SIGGRAPH Comput. Graph.* 15, 3 (Aug. 1981), 109–119. https://doi.org/10.1145/965161.806796

Feridun M. Celebi, Elizabeth S. Kim, Quan Wang, Carla A. Wall, and Frederick Shic. 2014. A Smooth Pursuit Calibration Technique. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA ’14)*. ACM, New York, NY, USA, 377–378. https://doi.org/10.1145/2578153.2583042

Dietlind Helene Cymek, Antje Christine Venjakob, Stefan Ruff, Otto Hans-Martin Lutz, Simon Hofmann, and Matthias Roetting. 2014. Entering PIN codes by smooth pursuit eye movements. *Journal of Eye Movement Research* 7, 4 (2014). https://bop.unibe.ch/index.php/JEMR/article/view/2384

Heiko Drewes, Alexander De Luca, and Albrecht Schmidt. 2007. Eye-gaze Interaction for Mobile Phones. In *Proceedings of the 4th International Conference on Mobile Technology, Applications, and Systems and the 1st International Symposium on Computer Human Interaction in Mobile Technology (Mobility ’07)*. ACM, New York, NY, USA, 364–371. https://doi.org/10.1145/1378063.1378122

Heiko Drewes, Mohamed Khamis, and Florian Alt. 2018. Smooth Pursuit Target Speeds and Trajectories. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia (MUM 2018)*. ACM, New York, NY, USA, 139–146. https://doi.org/10.1145/3282894.3282913

Heiko Drewes and Albrecht Schmidt. 2007. Interacting with the Computer Using Gaze Gestures. In *Proceedings of the 11th IFIP TC 13 International Conference on Human-computer Interaction - Volume Part II (INTERACT'07)*. Springer-Verlag, Berlin, Heidelberg, 475–488. http://dl.acm.org/citation.cfm?id=1778331.1778385

Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2015. Orbits: Gaze Interaction for Smart Watches Using Smooth Pursuit Eye Movements. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. ACM, New York, NY, USA, 457–466. https://doi.org/10.1145/2807442.2807499

David R Flatla, Carl Gutwin, Lennart E Nacke, Scott Bateman, and Regan L Mandryk. 2011. Calibration games: making calibration tasks enjoyable by adding motivating game elements. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*. ACM, 403–412.

Argenis Ramirez Gomez and Hans Gellersen. 2018. Smooth-i: Smart Re-calibration Using Smooth Pursuit Eye Movements. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications (ETRA '18)*. ACM, New York, NY, USA, Article 10, 5 pages. https://doi.org/10.1145/3204493.3204585

Elias Daniel Guestrin and Moshe Eizenman. 2008. Remote Point-of-gaze Estimation Requiring a Single-point Calibration for Applications with Infants. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications (ETRA '08)*. ACM, New York, NY, USA, 267–274. https://doi.org/10.1145/1344471.1344531

Rob Jacob and Sophie Stellmach. 2016. What You Look at is What You Get: Gaze-based User Interfaces. *interactions* 23, 5 (Aug. 2016), 62–65. https://doi.org/10.1145/2978577

Robert JK Jacob and Keith S Karn. 2003. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In *The mind's eye*. Elsevier, 573–605.

Mohamed Khamis, Florian Alt, and Andreas Bulling. 2016a. Challenges and Design Space of Gaze-enabled Public Displays. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (UbiComp '16)*. ACM, New York, NY, USA, 1736–1745. https://doi.org/10.1145/2968219.2968342

Mohamed Khamis, Ozan Saltuk, Alina Hang, Katharina Stolz, Andreas Bulling, and Florian Alt. 2016b. TextPursuits: Using Text for Pursuits-based Interaction and Calibration on Public Displays. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 274–285. https://doi.org/10.1145/2971648.2971679

Vassilios Krassanakis, Vassiliki Filippakopoulou, and Byron Nakos. 2016. Detection of moving point symbols on cartographic backgrounds. *Journal of Eye Movement Research* 9, 2 (2016).

Carlos Morimoto and Marcio R.M. Mimica. 2005. Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding* 98 (04 2005), 4–24. https://doi.org/10.1016/j.cviu.2004.07.010

Michaela Murauer, Michael Haslgrübler, and Alois Ferscha. 2018. Natural Pursuits for Eye Tracker Calibration. In *Proceedings of the 5th International Workshop on Sensor-based Activity Recognition and Interaction (iWOAR '18)*. ACM, New York, NY, USA, Article 3, 10 pages. https://doi.org/10.1145/3266157.3266207

Takehiko Ohno and Naoki Mukawa. 2004. A free-head, simple calibration, gaze tracking system that enables gaze-based interaction. In *Proceedings of the 2004 symposium on Eye tracking research & applications*. ACM, 115–122.

Ken Pfeuffer, Melodie Vidal, Jayson Turner, Andreas Bulling, and Hans Gellersen. 2013. Pursuit Calibration: Making Gaze Calibration Less Tedious and More Flexible. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. ACM, New York, NY, USA, 261–270. https://doi.org/10.1145/2501988.2501998

Argenis Ramirez-Gomez and Hans Gellersen. 2017. GazeBall: Leveraging Natural Gaze Behavior for Continuous Re-calibration in Gameplay. *Journal of Eye Movement Research* 10, 6 (2017).

Thiago Santini, Wolfgang Fuhl, and Enkelejda Kasneci. 2017. CalibMe: Fast and Unsupervised Eye Tracker Calibration for Gaze-Based Pervasive Human-Computer Interaction. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 2594–2605. https://doi.org/10.1145/3025453.3025950

Susan K. Schnipke and Marc W. Todd. 2000. Trials and Tribulations of Using an Eye-tracking System. In *CHI '00 Extended Abstracts on Human Factors in Computing Systems (CHI EA '00)*. ACM, New York, NY, USA, 273–274. https://doi.org/10.1145/633292.633452

Yusuke Sugano and Andreas Bulling. 2015. Self-Calibrating Head-Mounted Eye Trackers Using Egocentric Visual Saliency. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. ACM, New York, NY, USA, 363–372. https://doi.org/10.1145/2807442.2807445

Subarna Tripathi and Brian Guenter. 2017. A Statistical Approach to Continuous Self-Calibrating Eye Gaze Tracking for Head-Mounted Virtual Reality Systems. 862–870. https://doi.org/10.1109/WACV.2017.101

Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2013a. Pursuits: Spontaneous Interaction with Displays Based on Smooth Pursuit Eye Movement and Moving Targets. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13)*. ACM, New York, NY, USA, 439–448. https://doi.org/10.1145/2493432.2493477

Mélodie Vidal, Ken Pfeuffer, Andreas Bulling, and Hans W. Gellersen. 2013b. Pursuits: Eye-based Interaction with Moving Targets. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)*. ACM, New York, NY, USA, 3147–3150. https://doi.org/10.1145/2468356.2479632

Arantxa Villanueva and Rafael Cabeza. 2008. A novel gaze estimation system with one calibration point. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 38, 4 (2008), 1123–1138.

Arantxa Villanueva, Rafael Cabeza, and Sonia Porta. 2004. Eye Tracking System Model with Easy Calibration. In *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications (ETRA '04)*. ACM, New York, NY, USA, 55–55. https://doi.org/10.1145/968363.968372

Yanxia Zhang, Andreas Bulling, and Hans Gellersen. 2013. SideWays: A Gaze Interface for Spontaneous Interaction with Situated Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 851–860. https://doi.org/10.1145/2470654.2470775

Yanxia Zhang, Andreas Bulling, and Hans Gellersen. 2014a. Pupil-canthi-ratio: a calibration-free method for tracking horizontal gaze direction. In *Proc. of the 2014 International Working Conference on Advanced Visual Interfaces (AVI 14)* (2014-05-27). ACM, New York, NY, USA, 129–132. http://dx.doi.org/10.1145/2598153.2598186

Yanxia Zhang, Jörg Müller, Ming Ki Chong, Andreas Bulling, and Hans Gellersen. 2014b. GazeHorizon: Enabling Passers-by to Interact with Public Displays by Gaze. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*. ACM, New York, NY, USA, 559–563. https://doi.org/10.1145/2632048.2636071