

SacCalib: Reducing Calibration Distortion for Stationary Eye Trackers Using Saccadic Eye Movements

Michael Xuelin Huang
Max Planck Institute for Informatics
Saarland Informatics Campus
mhuang@mpi-inf.mpg.de

Andreas Bulling
Institute for Visualisation and Interactive Systems
University of Stuttgart
andreas.bulling@vis.uni-stuttgart.de

ABSTRACT

Recent methods to automatically calibrate stationary eye trackers were shown to effectively reduce inherent calibration distortion. However, these methods require additional information, such as mouse clicks or on-screen content. We propose the first method that only requires users' eye movements to reduce calibration distortion in the background while users naturally look at an interface. Our method exploits that calibration distortion makes straight saccade trajectories appear curved between the saccadic start and end points. We show that this curving effect is systematic and the result of distorted gaze projection plane. To mitigate calibration distortion, our method undistorts this plane by straightening saccade trajectories using image warping. We show that this approach improves over the common six-point calibration and is promising for reducing distortion. As such, it provides a non-intrusive solution to alleviating accuracy decrease of eye tracker during long-term use.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**;

KEYWORDS

Eye Tracking; Implicit Calibration; Eye Movements; Saccades

ACM Reference Format:

Michael Xuelin Huang and Andreas Bulling. 2019. SacCalib: Reducing Calibration Distortion for Stationary Eye Trackers Using Saccadic Eye Movements. In *Communication by Gaze Interaction (COGAIN @ ETRA'19)*, June 25–28, 2019, Denver, CO, USA. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3317956.3321553>

1 INTRODUCTION

Eye tracking is flourishing given recent advances in hardware and software [Huang et al. 2017; Zhang et al. 2018] as well as given increasing demands for mainstream applications, such as gaming or foveated rendering. To achieve high tracking accuracy, eye trackers need to be calibrated to each individual user prior to first use. During calibration, a gaze projection plane is estimated by asking users to fixate at predefined on-screen locations [Duchowski 2017]

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
COGAIN @ ETRA'19, June 25–28, 2019, Denver, CO, USA

© 2019 Association for Computing Machinery.
ACM ISBN 978-1-4503-6728-8/19/06...\$15.00
<https://doi.org/10.1145/3317956.3321553>

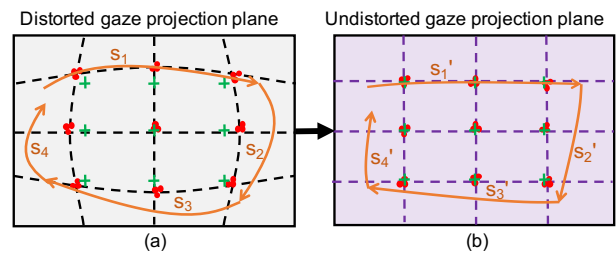


Figure 1: We correct the distorted gaze projection plane (black dash lines) to the undistorted gaze projection plane (purple dash lines) by minimizing the curvature of curved saccade trajectories (orange). By doing so, gaze points (red) are transformed closer to ground truth locations (green).

or to follow a moving dot [Pfeuffer et al. 2013]. While high accuracy is achieved right after calibration, significant decrease was demonstrated during use [Sugano and Bulling 2015], due to changes in users' head pose, relative position between screen and eye tracker, and other factors [Blignaut 2016]. We refer to the mapping from estimated gaze point onto ground truth as *calibration distortion*. Previous works proposed post-hoc correction [Špakov and Gizatdinova 2014], or to embrace it in the design of error-aware gaze interfaces [Barz et al. 2018]. However, these approaches only alleviate the symptoms and do not address the problem directly.

Another line of work introduced the idea of *self-calibration*, i.e. continuous recalibration in the background while the eye tracker is being used [Sugano and Bulling 2015]. While this approach was shown to be effective, current self-calibration methods either assume correlated, secondary user input, such as mouse clicks [Huang et al. 2016] or touch input [Zhang et al. 2018], or require information about on-screen content to compute saliency maps [Sugano and Bulling 2015]. While user input and gaze are often correlated [Huang et al. 2016; Sugano et al. 2015], this correlation is far from perfect and cannot be guaranteed. In addition, collecting sufficient and high-quality interaction data remains challenging. Similarly, while saliency maps can predict likely on-screen gaze locations [Sugano and Bulling 2015], the reliability of these predictions can be low and the computational cost of computing them can cause problems for real-time operation.

To address these limitations we propose *SacCalib* – the first calibration distortion reduction method that only requires information about a user's eye movements. That is, without the need for secondary user input or time-consuming computation of saliency maps for on-screen content. Saccade trajectories recorded by a calibrated eye tracker are nearly straight between the saccadic start and end

points, i.e. form a regular gaze projection plane (see Figure 1b). The key observation that our method builds on is that *without calibration straight saccade trajectories appear curved*, resulting in a *distorted gaze projection plane* (see Figure 1a). We see that this curving effect is systematic. That is, by observing multiple saccades performed between different on-screen locations, and by jointly minimizing saccade curvatures and thus undistorting the corresponding gaze projection plane, calibration distortion can be reduced.

A key challenge for our method is that saccadic eye movements, while straight in principle, can be curved. However, as suggested in [Godijn and Theeuwes 2004; Moehler and Fiehler 2014, 2015], natural saccade curvature can be reduced given sufficient preparation time for performing a saccade. We thus design an experiment paradigm to capture saccades under such condition and demonstrate the effectiveness of eye-only calibration distortion reduction. As such, our method opens up an exciting new avenue for eye tracker self-calibration and also paves the way for numerous practical gaze-aware HCI applications that do not require frequent, cumbersome, and time-consuming explicit eye tracker recalibration.

The contributions of our work are two-fold. First, we propose the first eye-only calibration distortion reduction technique based only on saccadic eye movements. In contrast to current self-calibration methods, our method neither requires secondary user input, such as mouse clicks, nor expensive processing of on-screen content. Second, we evaluate our method on a newly collected, 10-participant dataset of around 3,000 on-screen saccades, which we will release to the research community upon acceptance. Through this evaluation, we provide insight into the key issues of eye-only calibration.

2 RELATED WORK

Our work is informed by research on natural curvature of saccadic eye movements and prior work on eye tracker self-calibration.

2.1 Curvature of Saccadic Eye Movements

The reason for saccade curvature [Van der Stigchel 2010; Viviani et al. 1977] is still an open research question [Kruijne et al. 2014; Smit and Van Gisbergen 1990; Van der Stigchel et al. 2006]. Potential causes include oculomotor inhibition [Doyle and Walker 2001; Tipper et al. 1997], saccadic latency [McSorley et al. 2006], top-down selection processes [Van der Stigchel et al. 2006], and residual motor activity [Rizzolatti et al. 1987; Wang et al. 2011]. Oculomotor inhibition denotes competing saccade programs for the target and task-irrelevant distractor, which cause saccade deviation away from or toward the distractor [McPeck et al. 2003]. It may also be related to saccade attributes, such as saccade (early/late) stages [McSorley et al. 2006; Walker et al. 2006] or *saccadic latency* between saccade starting indicator and saccade onset [Ludwig and Gilchrist 2003; McSorley et al. 2006; Moehler and Fiehler 2015]. On the other hand, it was argued that saccade curvature may stem from top-down selection processes of the target [Van der Stigchel et al. 2006]. An unresolved competition among visual targets results and a clear goal-directed orienting may lead to different saccade curvatures. In addition, the second saccade in consecutive saccades could curve away from the initial fixation [Megardon et al. 2017]. This is regarded as residual motor activity [Wang et al. 2011].

It is worth noting that saccade curvature has been observed to decline or even vanish with an increase of *movement preparation time* preceding the saccades [Godijn and Theeuwes 2004; Moehler and Fiehler 2014, 2015]. This implies that sufficient time allows for the completion of top-down selection process among targets and thus reduces oculomotor inhibition and leads to straight saccades.

2.2 Eye Tracker Self-Calibration

Most commercial eye trackers relies on explicit calibration, which requires users to fixate on on-screen calibration points. For example, Tobii eye tracker applies a six-point calibration procedure that shows four points on screen corners and two in center. Although this procedure is usually short, a frequent re-calibration to maintain eye tracking accuracy is intrusive to users. Therefore, increasing research demands has risen for eye tracker implicit self-calibration.

Eye tracker self-calibration can be categorized into post-hoc correction [Barz et al. 2018; Špakov and Gizatdinova 2014], saliency-based [Chen and Ji 2015; Sugano and Bulling 2015; Sugano et al. 2013; Wang et al. 2016] and eye movement based approaches [Huang et al. 2016; Khamis et al. 2016; Papoutsaki et al. 2016; Pfeuffer et al. 2013; Sugano et al. 2015]. The first one conducts post-hoc correction of gaze estimation without modifying the gaze projection plane, while the other two aim to overcome calibration distortion by correcting the plane distortion as this study. Specifically, saliency-based method extracts *saliency map* of either screen image or user's ego-centric view and then maps eye features into image coordinate by assuming that the user is likely to look at the most salient locations.

Early works used bottom-up saliency maps that model the influence of low-level image attributes, such as edge, shape, and color [Itti et al. 1998; Koch and Ullman 1987], as well as the high-level image semantics [Huang et al. 2015], including objects [Xu et al. 2014], human faces [Sugano and Bulling 2015], gaze location of a person [Gorji and Clark 2017] or multiple persons [Fan et al. 2018]. More recent works investigated top-down saliency maps that account for goal-oriented or task-controlled visual attention and cognitive processes [Borji et al. 2012; Huang et al. 2018; Peters and Itti 2007]. However, calculating saliency maps for each camera frame is computational expensive and saliency maps can be highly inconsistent with users' actual visual attention [Judd et al. 2009].

In contrast, eye movement based approaches exploit secondary user input. Specifically, conventional approaches include fixation-based [Huang et al. 2016; Papoutsaki et al. 2016; Sugano et al. 2015] and pursuit-based [Khamis et al. 2016; Pfeuffer et al. 2013]. The general assumption of fixation-based methods is that the user looks at the interaction location, such as locations of mouse-clicks [Huang et al. 2016; Papoutsaki et al. 2016; Sugano et al. 2015], mouse movements [Huang et al. 2012; Papoutsaki et al. 2016], and key presses [Huang et al. 2016]. In contrast, pursuit-based calibration relies on the movement correlation between visual stimuli and eye gaze, and the moving stimuli can be a specific cursor [Pfeuffer et al. 2013], natural texts [Khamis et al. 2016], or an object in games [Tripathi and Guenter 2017]. Despite the close link between eye movement and user input, eye movement-based calibration is limited by input sparsity in real use, and pursuit-based calibration requires dynamic interfaces. Therefore, eye-only calibration that

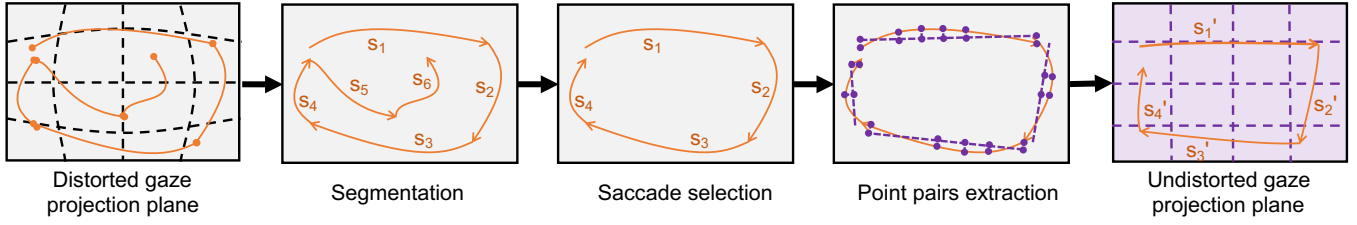


Figure 2: Our method segments the saccades, selects those with suitable properties (see main text for details), and extracts point pairs on the curved saccade trajectories and its straight counterpart. It then corrects the distorted gaze projection plane to the undistorted projection plane by jointly minimizing saccade curvatures, thereby effectively reducing calibration distortion.

does not require visual scene or user input can be highly beneficial to addressing the limitations of conventional calibration techniques.

3 REDUCING DISTORTION VIA WARPING

To correct the distorted gaze projection plane, we first identify and segment saccades using the velocity threshold method (I-VT) [Salvucci and Goldberg 2000] following common practice [Arabadzhiyska et al. 2017]. We then select suitable saccades and extract all pairs of gaze points along the curved and straight saccade trajectories. These point pairs are input to an image warping technique that uses moving least squares [Schaefer et al. 2006] to correct curved saccade trajectories to straight. As a result, the distorted gaze projection plane is undistorted. Figure 2 shows the method overview.

3.1 Extracting Point Pairs for Warping

The basis of our eye-only calibration method lies in the rectification of the curved saccade trajectories to (assumed) straight trajectories. However, the solution to correction is not unique. Figure 3 shows three potential projections with different offsets from the saccade endpoints. By maintaining the saccade direction pointing from saccade start to its end, we can a) project the curved saccade trajectory to the straight line that connects its two endpoints; we can also b) shift it to the peak of the distorted saccade, so that they just touch each other, or c) shift it to the middle toward the peak. We adopt the projection with shift to the middle in our method. Let $\langle p_1, \dots, p_m \rangle$ and $\langle p'_1, \dots, p'_m \rangle$ be the gaze points on all the distorted saccades, and their projection counterparts on the corrected saccades, respectively. The point pairs that control the warping of the gaze project plane can be represented by $\langle \{p_1, p'_1\}, \dots, \{p_m, p'_m\} \rangle$.

3.2 Undistorting the Gaze Projection Plane

Inspired by image warping technique [Schaefer et al. 2006], we undistort the gaze projection plane by minimizing the distance between gaze point pairs on curved and straight saccade trajectories.

There are three desired properties of undistorting the gaze projection plane. First, it should reduce the gap between gaze point pairs. Second, it should produce smooth deformations, i.e. the area among different gaze point pairs should be smooth. Third, it should preserve the original relative geometry, as e.g. a point on the left side of a saccade is expected to stay on the left after warping.

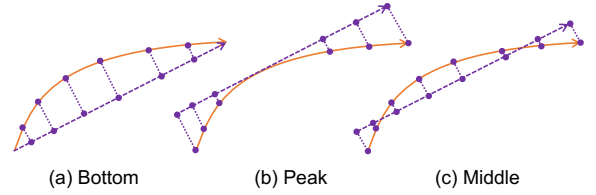


Figure 3: Curved saccade trajectory (orange curve) is corrected to straight saccade (purple dash line) by (a) projecting to the line that connects two endpoints; or (b) with an additional shift to the peak of the curved saccade; or (c) to the middle toward the peak. Purple dots indicate gaze points.

To this end, given a point v , we solve for the affine transformation $l_v(x)$ that minimizes

$$\sum_i^m w_i |l_v(p_i) - p'_i|^2 \tag{1}$$

where p_i and p'_i are the point pair that controls warping and w_i are the weights that control the impact of each point pair on the transformation of point v . Intuitively, the weights should be inversely related to the distance from the input points to achieve the smoothness of transformation, thus we define it as $w_i = 1/|p_i - v|$.

As pointed out in [Schaefer et al. 2006], the affine transformation $l_v(x)$ should consist of a linear transformation and a translation, but the translation component can be substituted by referring to the weighted centroids of the point pairs. That is, Equation 1 can be rewritten in term of the linear matrix M .

$$\sum_i^m w_i |\hat{p}_i M - \hat{p}'_i|^2 \tag{2}$$

where $\hat{p}_i = p_i - \sum_i w_i p_i / \sum_i w_i$ and $\hat{p}'_i = p'_i - \sum_i w_i p'_i / \sum_i w_i$ are the normalized point pair by their weighted centroids, respectively. Depending on the form of matrix M , we can fine-control the transformation characteristics. Specifically, using the general form of matrix M results in a fully affine transformation, which contains non-uniform scaling and shear. Restricting matrix M to a similarity transformation that only includes translation, rotation, and uniform scaling better preserves angles on the original plane. Further restricting matrix M to a rigid transformation that excludes scaling can maintain the relative geometry after warping. We therefore use

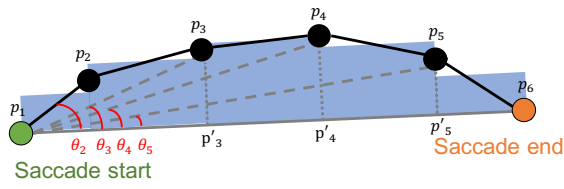


Figure 4: Illustration of the area- and direction-based measures for saccade curvature. The former one computed the area covered by the curved saccade trajectory, while the latter is the average angle of gaze points w.r.t. the endpoints.

a rigid transformation form of matrix M . Please refer to [Schaefer et al. 2006] for more information.

To speed-up the computation, we approximate the full plane with a fine-grained grid with a quad size of 25 pixels and only apply the deformation to each quad vertex. We then perform bilinear interpolation to fill each pixel in the quads.

Due to the potentially contradictory warping controlled by different point pairs, the undistorted projection plane may suffer from undesirable fold-back, where a point on one side of a line may be mapped to the other side. Therefore, we apply a post-hoc spatial smoothing on the resulting transformation, by using a normalized box filter with a blurring kernel size of 5 on the grid data.

3.3 Selecting Saccades

Although the above method can improve gaze accuracy by correcting the distorted saccades, in practice not all saccades are purely deformed by the distorted gaze projection plane. Instead, saccadic eye movements can be noisy, sometimes too short, and mixed with jittering [Farmer and Sidorowich 1991]. The ideal saccade candidates for our method are those that are straight by nature. In addition, they should be long enough to impose a valid effect on the plane correction. Therefore, we filter saccades before applying only the suitable candidates to undistort the gaze projection plane. Specifically, we construct a random forest classifier that selects saccades based on multiple attributes. We start with the discussion about the curvature measure, given its close link to our core idea.

3.3.1 Measure of saccade curvature. Measures of saccade curvature can be direction-based, distance-based, area-based, or curve fitting-based [Tudge et al. 2017; Van der Stigchel et al. 2006]. These measures can be computed with respect to the location of the target or the endpoint of saccade. As we aim for eye-only calibration, where the actual target is agnostic to the method, the current study focuses on the endpoint-based measures. We compute the area-based curvature as well as the direction-based curvature, as suggested in [Tudge et al. 2017]. Specifically, the *area-based curvature* measures the area between the saccade trajectory and the straight line from saccade start to saccade end. As shown in Figure 4, let p_i in $\langle p_1, \dots, p_m \rangle$ denote the i -th point on a saccade with m points, the area-based curvature, *CurvatureArea*, is computed by

$$CurvatureArea = \sum_{i=2}^m \frac{|p_i p'_i| + |p_{i-1} p'_{i-1}|}{2} |p'_i p'_{i-1}| \quad (3)$$

Table 1: We extract 10 saccade attributes. The number of attributes in each type is shown in the parenthesis.

Type	Measure of the attributes
Curvature (2)	<i>CurvatureArea</i> is an area-based measure and <i>CurvatureAngle</i> is a direction-based measure according to Equation 3 and Equation 4, respectively.
Amplitude (2)	<i>Amplitude</i> measures the direct distance between saccade start and end, while <i>Length</i> measures the sum distance between all consecutive gaze points on a saccade.
Orientation (2)	<i>Direction</i> denotes the saccade angle with respect to the horizontal line; <i>Turn</i> presents the number of large direction change ($>90^\circ$).
Velocity (2)	<i>Velocity</i> measures the overall velocity of all points on a saccade, and <i>VelocityMax</i> delineates the peak velocity.
Timing (2)	<i>Latency</i> measures the interval between the vanishing of the last target and the saccade start, and <i>Duration</i> measure the saccade duration.

The *direction-based curvature* denotes the average angles formed by lines from saccade start to each gaze point on the saccade, with respect to the straight line connecting two saccade endpoints. Let θ_i be the i -th angle, i.e. $\angle p_i p_1 p_m$, the direction-based curvature, *CurvatureAngle*, is quantified by

$$CurvatureAngle = \frac{1}{m} \sum_{i=2}^{m-1} \theta_i \quad (4)$$

3.3.2 Identifying suitable saccades using a data-driven approach. Apart from saccade curvature, we also consider other types of attributes, including amplitude, orientation, velocity, and timing (see Table 1). The reason for these attributes stems from findings about natural saccade curvature in human vision research. For instance, vertical and oblique saccades were found to be more curved than horizontal ones [Bahill and Stark 1975; Kowler 2011; Yarbus 1967]. Similarly, saccade amplitude can be pertinent [Van Opstal and Van Gisbergen 1987]. Besides, saccadic latency was found to be related to saccade curvature [Ludwig and Gilchrist 2003; McSorley et al. 2006; Moehler and Fiehler 2015]. As such, saccade orientation, amplitude, and timing can be good indicators for its natural curvature. In addition, given the close link between saccade curvature and its velocity profile, we also include the velocity attributes.

We extract and use these attributes as input to a random forest classifier to predict the suitability for correction of each saccade. We set the number of attributes in each tree to 5, the number of trees in the forest to 20, the maximum depth of the tree to 50 to reduce the risk of overfitting due to our medium-size training data.

3.3.3 Acquiring training data. To obtain the training samples, we perform a leave-one-saccade-out procedure on the session-based recording data. That is, we start with using all the saccades for plane correction. We then iteratively leave out one random saccade if the removal improves the overall accuracy. We stop the procedure

until no more improvement can be achieved. This gives us a *suitable saccade subset* that contributes to the plane correction.

In our setting, the remaining saccades from different sessions take approximately 10% to 30% of the original saccades. To create a balanced training set for the random forest classifier, we perform a random downsampling on the major class (i.e. *unsuitable saccade subset*). Finally, we group the suitable and unsuitable subsets of saccades from different sessions and participants for training.

3.4 Segmenting Saccadic Eye Movements

Eye tracking data mainly contains saccades, fixations, and blinks. We follow previous practice to segment saccades, according to the velocity profile of the eye movements [Arabadzhiyska et al. 2017].

3.4.1 Fixing the noisy eye tracking data. Raw eye tracking can be noisy, due to eye blinks, poor tracking quality, motion blur, and infrared reflection on glasses. We first remove high-frequency jitter. That is, medium ($\sim 1^\circ$) jerk-like eye movements at an abnormal frequency of around 100 Hz that occur frequently near the screen boundary. We then use linear interpolation to fill missing data of short durations (< 50 ms). Such data loss is likely a result from visual noise or tracking failure, for normal eye blink duration is around 100-140 ms [Schiffman 1990]. We finally apply a low-pass filter [Farmer and Sidorowich 1991] to remove high-frequency noise.

3.4.2 Segmenting saccades. After the above preprocessing, we apply the I-VT method [Salvucci and Goldberg 2000] to segment saccadic eye movements. Specifically, we define three velocity thresholds as in [Arabadzhiyska et al. 2017; Dorr et al. 2010]: a detection threshold V_d ($100^\circ/s$), a starting velocity threshold V_a ($60^\circ/s$), and a final velocity threshold V_f ($60^\circ/s$). The *detection threshold* is used to identify the first gaze point whose velocity exceeds V_d and we refer to it as the *detection point*. This threshold identifies a gaze point on a saccade with a safe margin. Since the actual saccade start is earlier than the detection point, we scan backward from the detection point and look for the saccade start, where velocity begins to exceed V_a . Similarly, we seek forward for the gaze point whose velocity begins to drop beyond V_f and refer to it as the saccade end. The values of these threshold parameters are in good agreement with prior studies [Arabadzhiyska et al. 2017; Boghen et al. 1974].

4 COLLECTING A SACCADE DATASET

To study the feasibility and effectiveness of the proposed eye-only calibration method we collected a 10-participant dataset of saccades with different amplitudes and directions.

4.1 Experiment Design

Since the proposed method assumes straight saccades for the correction of the gaze projection plane, we designed our data collection procedure so as to capture natural straight saccades. Specifically, we used a shrinking circle shown on a black screen to direct the saccadic eye movement of the participants. As shown in Figure 5, the current visual target is shown initially in white with a red dot in its center, and it gradually turns green in one second and then shrinks and disappears in another second. Once the current (green) target starts to shrink, the next (white) target shows up. When the current target shrinks to vanish, participants should make a

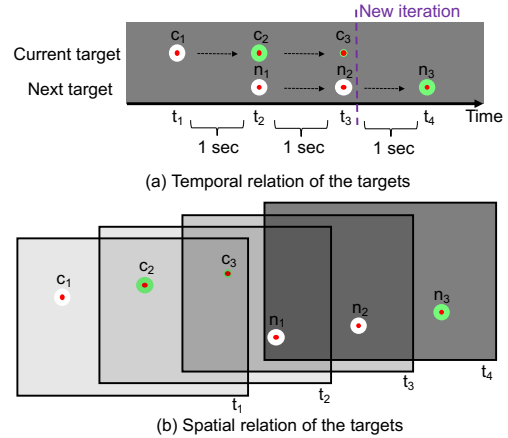


Figure 5: The (a) temporal and (b) spatial relation of the current and the next visual targets. The current target (in green) shrinks and disappears, but once it starts to shrink the next target (in white) is shown in another random location defined by a five-by-five grid that evenly covers the screen.

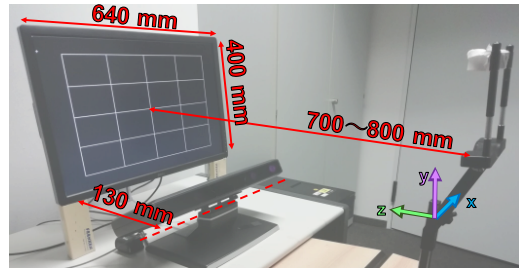


Figure 6: Experiment setup. A 30-inch monitor, a Tobii TX300 eye tracker and a chin rest were used for recording. We adjusted the location of the table, where the chin rest was mounted on, across sessions to simulate head pose variation after initial calibration.

saccade to the next target. To minimize the impact of oculomotor inhibition and top-down selection processes, this study focuses on the scenario with no distraction for saccades.

Visual targets were shown in random locations on a five-by-five grid spread evenly across the screen, resulting in saccades with diverse directions and amplitudes. This was inspired by prior studies showing that horizontal saccades are more likely to be straight [Yarbus 1967], while vertical and oblique saccades generally appear curved [Kowler 2011]. As suggested by Moehler and Fiehler [Moehler and Fiehler 2015], the saccade curvature effect vanishes after one second saccade preparation time. Therefore, one second before the current visual target disappeared, the next target already appeared in another location on the screen. We encouraged participants to locate that next target using peripheral vision, i.e. to shift their covert attention toward it once it appeared. To help users discriminate between targets, the current target was always shown in green while the next one was shown in white.

4.2 Apparatus

For data recording, we used a stationary Tobii TX300 eye tracker sampling at 300 Hz. We used a chin rest to constrain head pose variation (see Figure 6). The experiment interface was shown full-screen on a 30-inch (640x400 mm; $\sim 50^\circ$, resolution of 2560x1600 pixels), which was placed approximate 750 mm away from the chin rest. The diameter of the circular stimuli before shrinking was 40 pixels ($\sim 0.8^\circ$), corresponding to the best reported precision of the eye tracker. We changed the chin rest location across sessions to maximize the inter-session differences caused by head pose and study such impact on our performance. This is because there should be an one-to-one mapping between head pose and the undistorted gaze projection plane, and the current study focuses on the plane correction without considering dynamic head pose changes.

4.3 Participants

We recruited 11 participants (three female; average age: 28.3), among which three are Indian, five are Asian, and three are Caucasian. Four of them wore glasses. A close scrutiny reveals that the data of one participant had a poor condition, which presented an extra bad precision (i.e. a severer jittering/dispersion during fixation) and low accuracy (i.e. an obvious deviation from ground truth), and contained a large proportion of invalid tracking frames. We therefore exclude this participant. This gives us a 10-participant dataset.

4.4 Procedure

Participants were first introduced to the experiment interface and allowed to familiarize themselves with the interface for around a minute. Participants then performed a standard six-point calibration using the Tobii eye tracker interface, followed by three sessions of recordings, each of which lasted for about five minutes. The first session directly followed the eye tracker calibration, while the second and the third sessions were conducted with a different amount of head position change. This is to simulate the impact of head pose change after initial calibration. To this end, each time we adjusted the position of the chin rest randomly in x-, y-, and z-direction. More specifically, the range of the first position change was in a medium degree (around 40 mm), and the second change was in a large degree (around 80 mm; reaching the boundary of the valid range of the eye tracker). After each adjustment of the chin rest, participants were also allowed to tune the height and the position of the chair for the most comfortable condition.

In each session, the visual target traversed the five-by-five grid in a random order for five times, which resulted in around a hundred saccadic eye movements per participant. In order words, our data in total contains 30 recording sessions and approximately 3,000 saccadic eye movements. Between sessions participants were encouraged to rest, walk around, and look outside the window. These breaks lasted for at least one minute and were extended up to around five minutes if requested by the participants. On average, the entire experiment recording took about 20 minutes per participant.

4.5 Distortion across Pose Variations

Two widely used metrics for eye tracking performance are precision and accuracy [Duchowski 2017; Holmqvist et al. 2011]. *Precision* represents the deviation among gaze points of one fixation from their

centroid, while *accuracy* denotes the average distance from gaze points to the ground truth location. Please note that our method aims to improve the eye tracking accuracy by transforming the deviated gaze points toward the correct locations. In other words, this is to amend the undermined accuracy (rather than precision) caused by the poor quality of initial calibration or the changes of screen-tracker geometry and head pose.

As we use visual targets shown in a five-by-five grid to direct saccadic eye movements, the ground truth location of fixations was at the corresponding grid vertex. To measure accuracy, we calculate average Euclidean distance between each vertex and gaze points of all fixations that correspond to the vertex over one session.

We see that eye tracking accuracy generally decreases as the increase of head pose variation from initial calibration position. More specifically, the overall Euclidean distance between the gaze points and ground truth at initial calibration pose is $1.07 \pm 0.65^\circ$, and those with small and large head pose variation are $1.17 \pm 0.64^\circ$ and $1.18 \pm 0.70^\circ$, respectively.

5 EXPERIMENTAL EVALUATION

This section evaluates the effectiveness of the proposed eye-only calibration for reducing calibration distortion. We aim to answer three key questions pertinent to this study: 1) Can it improve eye tracking accuracy with and without head pose variation after initial calibration and across participants? 2) Is saccade selection necessary and what are the important attributes? 3) Which is the optimal projection method to correct distorted saccades?

In the following experimental evaluation, we present the results in a leave-one-participant-out paradigm. That is, each time we trained a random forest classifier on the data of nine participants and tested the undistortion effect on the left out participant. We repeated this process for 10 times and the final performance is the average result over all the iterations.

5.1 Improvement over Initial Calibration

To answer the first question, we look into the improvement over initial calibration for overall participants. Figure 7 (left) shows the improvement across head pose variations. Bars with different colors indicate plane correction based on different number (50, 100, and all) of saccades as well as saccades selected by data-driven approach (random forest). The black error bar presents standard error. For simplicity, we refer to initial calibration with none head pose variation as "None", with a small variation as "Small" and with a large variation as "Large" in the following texts. Most importantly, our method based on data-driven saccade selection produces a consistent improvement over initial calibration with and without head pose variations. Overall, we achieve a 2.5% improvement, pushing accuracy from 1.79° to 1.75° (equivalent to 36.3% of average accuracy difference between "None" and "Large" without correction).

In contrast, undistorting gaze projection plane using 50 random saccades consistently decreases the accuracy, compared to that of using initial calibration. Increasing the number of saccades to 100 improves the accuracy for "Small" and "Large", but still decreases in "None". Although further including all saccades (up to 300) improves the accuracy in both "None" and "Small" by a marked margin (2.1%), it fails to improve the accuracy in "None" likewise. This is probably

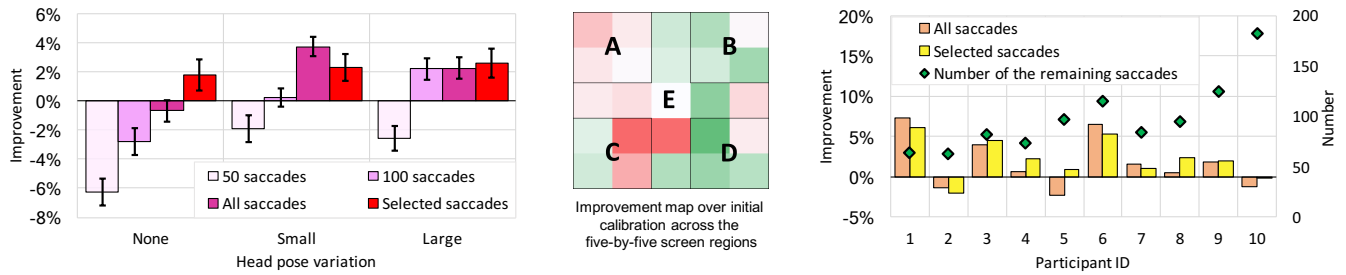


Figure 7: (Left) Improvement compared to initial calibration for each participant. The bars indicate the improvement when using all vs selected saccades. The green diamonds show the number of remaining saccades from data-driven saccade selection. With this selection, our method achieves consistent improvements for most participants. (Middle) Improvement map over initial calibration. Green indicates positive while red negative. Improvement of our method varies across screen regions probably due to data skewness of suitable saccades. (Right) Improvement over initial calibration across three positions while using 50, 100, and all saccades and data-driven selected saccades for plane correction. The error bars show standard error. Data-driven saccade selection yields a stable improvement over all participants. The negative value indicates a accuracy decrease.

because initial calibration without pose variation ($1.07 \pm 0.65^\circ$) is highly accurate and thus relatively hard to improve.

Interestingly, our improvement over initial calibration varies across screen regions. Since we used visual stimulus in a five-by-five grid during data collection, we visualize the improvement map accordingly (see Figure 7 middle). Green denotes accuracy increase and red decrease. Our method performs well to reduce distortion for overall participants. More specifically, it can significantly improve over initial calibration in three-fourths of the screen area ($B+C+D+E: p=0.028$ or $A+B+D+E: p=0.033$).

Inspecting improvement on individual participant (see Figure 7 right) suggests that our method can achieve improvement for majority (80%) participants. Further, we can reach approximately 5% improvement for almost one third of participants in overall scenarios across head pose variations. Importantly, data-driven approach for saccade selection is beneficial to plane correction for 70% participants. Interestingly, for a clear majority participants (70%), the number of remaining saccades after data-driven selection is between

50 to 100. However, a random selection with a similar number of saccades fails to achieve equivalent accuracy (see Figure 7 Left), implying that correct saccade selection is essential to our method.

5.2 Further Insight into Saccade Selection

To understand the difference between suitable and unsuitable saccades for the correction of gaze projection plane, we plot the probability mass functions of each saccade attribute (see Figure 8). In general, the probabilities of suitable and unsuitable saccades still peak at different values for some attributes, such as *Velocity* and *VelocityMax*, suggesting that suitable saccades tend to have a higher velocity as well as maximum velocity. In addition, the chance of saccade suitability varies in specific value range of some attributes. For example, it is likely to be a suitable saccade with a large *Length* or *Amplitude*, or a small positive *Latency*. In contrast, it is likely to be an unsuitable saccade with less than 50 ms duration. That the probabilities of suitable and unsuitable saccades

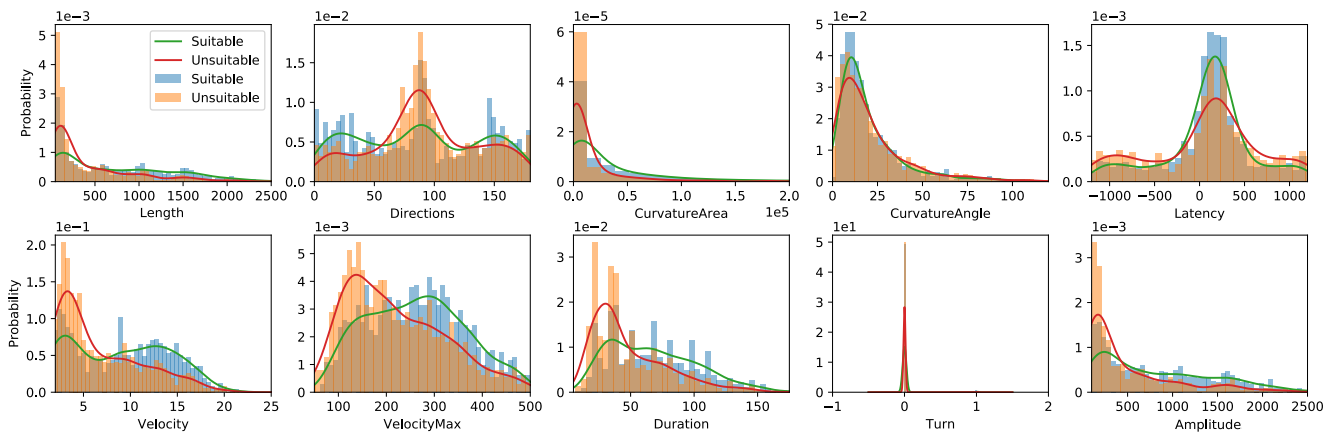


Figure 8: Probabilities mass functions of attributes of suitable and unsuitable saccades for the correction of gaze projection plane. Probability difference between suitable and unsuitable saccades exists in a certain value range of specific attributes.

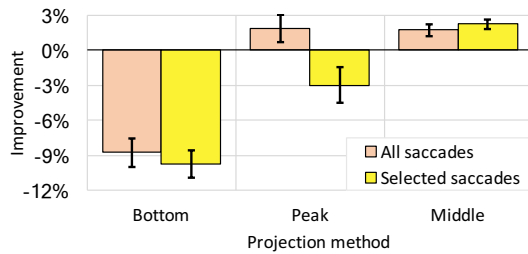


Figure 9: Performance comparison of different potential saccade projections. Projection to the middle between the saccade peak and bottom gives a consistent improvement over initial calibration using all and selected saccades.

are overlapped, motivates us to apply the data-driven approach to identifying saccade candidates for plane correction.

We also conducted the attribute importance analysis of our random forest classifiers through mean decrease impurity. In good agreement with the previous discussion, *Length* (ranks 2nd) and *Amplitude* (3rd) rank high in attribute importance. Most surprisingly, *CurvatureArea* rank first, suggesting that *CurvatureArea* can provide important complementary information for saccades selection, though itself alone is not informative enough (see Figure 8). This further indicates the need of data-driven saccade selection.

5.3 Effect of Saccade Projection Methods

We discuss three different potential projections to correct the distorted saccades in the method section. We hereby evaluate the effectiveness of these projections. Figure 9 shows the performance comparison of projecting the curve saccade to the line connecting two endpoints (Bottom) and its shifted counterpart to the saccade peak (Peak) and middle (Middle) between bottom and peak.

In general, projecting saccades to the middle between the saccade peak and bottom as used in our method is most promising. It achieves improvements for both using all and selected saccades. In contrast, projecting to peak achieves improvement only with all saccades, while projecting to bottom decreases initial calibration accuracy regardless of using all or selected saccades.

Most importantly, this result suggests that projection to the middle is a stable solution to correcting the distortion of gaze projection plane. It also points out the significant impact of different projection methods on the correction effect. As such, investigating alternative projection methods can be of great interest in future.

6 DISCUSSION

In this work we presented the first eye-only calibration method to reduce calibration distortion without user input or expensive processing of on-screen content. Specifically, we proposed to undistort gaze projection plane. As the first method of its kind, it was demonstrated to be effective. Further, we shed lights on two pertinent and critical issues: saccade selection and saccade projection method, i.e. what and how to undistort. These problems have been shown to be closely related to performance. As such, we believe this study represents a first important step towards eye-only calibration.

The results we achieved are encouraging. The proposed method is able to improve eye tracking accuracy directly after initial calibration, where the accuracy is relatively high and difficult to improve. Moreover, it can effectively reduce calibration distortion under small and large head pose variations after initial calibration. These improvements were consistent for most of our participants. For evaluation purposes of eye-only calibration, we collected a novel saccadic eye movement dataset. We believe the dataset will be beneficial to this new line of studies, and thus decided to release it upon acceptance and continue extending this dataset.

In practice, the proposed method has significant potential as a low-cost, non-intrusive, and privacy-preserving solution to reducing calibration distortion of stationary eye trackers. First, our method does not rely on additional information, such as computational expensive saliency map. Second, it allows for implicit calibration while users naturally look at an interface. Third, unlike previous approaches to eye tracker self-calibration, it does not require any additional user input or potentially privacy-sensitive information on on-screen content. These properties are valuable for practical real-time gaze-based interfaces.

We also identified a number of interesting directions for future work. First, due to the unbalance of ocular dominance [Chaurasia and Mathur 1976], the distortion of the gaze projection plane may differ for the left and the right eye. While in the current study we only investigated the data of the left eye, it will be interesting to study the relation between binocular coordination and the impact on the corresponding gaze projection planes. Second, in our experiment we ensured to capture saccades that are straight by nature by giving participants sufficient preparation time. To improve practicality of the approach, future work could investigate saccade identification using microsaccades [Engbert and Kliegl 2003] or on how to ensure sufficient preparation time in user interface design. Finally, as the first work in this area of research, our paper lays important foundations for future work but there is, of course, room for general performance improvements of the method itself. For example, we plan to account for the resulting gaze projection plane in saccade selection, which can maintain a good on-screen distribution of the selected saccades and thus a better warping performance.

7 CONCLUSION

In this work we proposed the first calibration method using saccadic eye movements that neither requires additional user input or expensive processing of on-screen content. We demonstrated its potential in reducing calibration distortion on a new saccade dataset and compared its performance to initial calibration with and without head pose variations. As such, the method provides a low-cost, non-intrusive, and privacy-preserving solution to reduce calibration distortion. We also identified two key challenges for eye-only calibration, namely saccade selection and saccade projection. While further research is required to make the approach practically usable, our results are promising and pave the way for a novel line of work on eye-only calibration for stationary eye trackers.

ACKNOWLEDGMENTS

This work was funded in part by the Cluster of Excellence on Multimodal Computing and Interaction at Saarland University, Germany.

REFERENCES

- Elena Arabadzhiyska, Okan Tarhan Tursun, Karol Myszkowski, Hans-Peter Seidel, and Piotr Didyk. 2017. Saccade landing position prediction for gaze-contingent rendering. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 50.
- A Terry Bahill and Lawrence Stark. 1975. Neurological control of horizontal and vertical components of oblique saccadic eye movements. *Mathematical Biosciences* 17, 3–4 (1975), 287–298.
- Michael Barz, Florian Daiber, Daniel Sonntag, and Andreas Bulling. 2018. Error-Aware Gaze-Based Interfaces for Robust Mobile Gaze Interaction. In *Proc. International Symposium on Eye Tracking Research and Applications (ETRA)*. 24:1–24:10. <https://doi.org/10.1145/3204493.3204536>
- Pieter Blignaut. 2016. Idiosyncratic feature-based gaze mapping. *JOURNAL OF EYE MOVEMENT RESEARCH* 9, 3 (2016).
- D Boghen, BT Troost, RB Daroff, LF Dell'Osso, and JE Birkett. 1974. Velocity characteristics of normal human saccades. *Investigative Ophthalmology & Visual Science* 13, 8 (1974), 619–623.
- Ali Borji, Dicky N Sihite, and Laurent Itti. 2012. Probabilistic learning of task-specific visual attention. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 470–477.
- BD Chaurasia and BBL Mathur. 1976. Eyedness. *Cells Tissues Organs* 96, 2 (1976), 301–305.
- Jixu Chen and Qiang Ji. 2015. A probabilistic approach to online eye gaze tracking without explicit personal calibration. *IEEE Transactions on Image Processing* 24, 3 (2015), 1076–1086.
- Michael Dorr, Thomas Martinetz, Karl R Gegenfurtner, and Erhardt Barth. 2010. Variability of eye movements when viewing dynamic natural scenes. *Journal of vision* 10, 10 (2010), 28–28.
- Melanie Doyle and Robin Walker. 2001. Curved saccade trajectories: Voluntary and reflexive saccades curve away from irrelevant distractors. *Experimental Brain Research* 139, 3 (2001), 333–344.
- Andrew T Duchowski. 2017. *Eye Tracking Methodology: Theory and Practice*. (2017).
- Ralf Engbert and Reinhold Kliegl. 2003. Microsaccades uncover the orientation of covert attention. *Vision research* 43, 9 (2003), 1035–1045.
- Lifeng Fan, Yixin Chen, Ping Wei, Wenguan Wang, and Song-Chun Zhu. 2018. Inferring Shared Attention in Social Scene Videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 6460–6468.
- J Doyne Farmer and John J Sidorowich. 1991. Optimal shadowing and noise reduction. *Physica D: Nonlinear Phenomena* 47, 3 (1991), 373–392.
- Richard Godijn and Jan Theeuwes. 2004. The relationship between inhibition of return and saccade trajectory deviations. *Journal of Experimental Psychology: Human Perception and Performance* 30, 3 (2004), 538.
- Siavash Gorji and James J Clark. 2017. Attentional push: A deep convolutional network for augmenting image saliency with shared attention modeling in social scenes. In *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*, Vol. 2. IEEE, 5.
- Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer. 2011. *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.
- Jeff Huang, Ryan White, and Georg Buscher. 2012. User see, user point: gaze and cursor alignment in web search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1341–1350.
- Michael Xuelin Huang, Tiffany CK Kwok, Grace Ngai, Stephen CF Chan, and Hong Va Leong. 2016. Building a personalized, auto-calibrating eye tracker from user interactions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 5169–5179.
- Michael Xuelin Huang, Jiajia Li, Grace Ngai, and Hong Va Leong. 2017. Screenglint: Practical, in-situ gaze estimation on smartphones. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2546–2557.
- Xun Huang, Chengyao Shen, Xavier Boix, and Qi Zhao. 2015. Salicon: Reducing the semantic gap in saliency prediction by adapting deep neural networks. In *Proceedings of the IEEE International Conference on Computer Vision*. 262–270.
- Yifei Huang, Minjie Cai, Zhenqiang Li, and Yoichi Sato. 2018. Predicting Gaze in Egocentric Video by Learning Task-dependent Attention Transition. In *European Conference on Computer Vision*.
- Laurent Itti, Christof Koch, and Ernst Niebur. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence* 20, 11 (1998), 1254–1259.
- Tilke Judd, Krista Ehinger, Frédo Durand, and Antonio Torralba. 2009. Learning to predict where humans look. In *Computer Vision, 2009 IEEE 12th international conference on*. IEEE, 2106–2113.
- Mohamed Khamis, Ozan Saltuk, Alina Hang, Katharina Stolz, Andreas Bulling, and Florian Alt. 2016. 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*. ACM, 274–285.
- Christof Koch and Shimon Ullman. 1987. Shifts in selective visual attention: towards the underlying neural circuitry. In *Matters of intelligence*. Springer, 115–141.
- Eileen Kowler. 2011. Eye movements: The past 25 years. *Vision research* 51, 13 (2011), 1457–1483.
- Wouter Kruijze, Stefan Van der Stigchel, and Martijn Meeter. 2014. A model of curved saccade trajectories: Spike rate adaptation in the brainstem as the cause of deviation away. *Brain and cognition* 85 (2014), 259–270.
- Casimir JH Ludwig and Iain D Gilchrist. 2003. Target similarity affects saccade curvature away from irrelevant onsets. *Experimental Brain Research* 152, 1 (2003), 60–69.
- Robert M McPeck, Jae H Han, and Edward L Keller. 2003. Competition between saccade goals in the superior colliculus produces saccade curvature. *Journal of Neurophysiology* 89, 5 (2003), 2577–2590.
- Eugene McSorley, Patrick Haggard, and Robin Walker. 2006. Time course of oculomotor inhibition revealed by saccade trajectory modulation. *Journal of Neurophysiology* 96, 3 (2006), 1420–1424.
- Geoffrey Megardon, Casimir Ludwig, and Petroc Sumner. 2017. Trajectory curvature in saccade sequences: spatiotopic influences vs. residual motor activity. *Journal of Neurophysiology* 118, 2 (2017), 1310–1320.
- Tobias Moehler and Katja Fiehler. 2014. Effects of spatial congruency on saccade and visual discrimination performance in a dual-task paradigm. *Vision research* 105 (2014), 100–111.
- Tobias Moehler and Katja Fiehler. 2015. The influence of spatial congruency and movement preparation time on saccade curvature in simultaneous and sequential dual-tasks. *Vision research* 116 (2015), 25–35.
- Alexandra Papoutsaki, Patsorn Sangkloy, James Laskey, Nediya Daskalova, Jeff Huang, and James Hays. 2016. Webgazer: Scalable webcam tracking using user interactions. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence-IJCAI 2016*.
- Robert J Peters and Laurent Itti. 2007. Beyond bottom-up: Incorporating task-dependent influences into a computational model of spatial attention. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 1–8.
- 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*. ACM, 261–270.
- Giacomo Rizzolatti, Lucia Riggio, Isabella Dascola, and Carlo Umiltà. 1987. Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention. *Neuropsychologia* 25, 1 (1987), 31–40.
- Dario D Salvucci and Joseph H Goldberg. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*. ACM, 71–78.
- Scott Schaefer, Travis McPhail, and Joe Warren. 2006. Image deformation using moving least squares. In *ACM transactions on graphics (TOG)*, Vol. 25. ACM, 533–540.
- Harvey Richard Schiffman. 1990. *Sensation and perception: An integrated approach*. Oxford, England: John Wiley & Sons.
- AC Smit and JAM Van Gisbergen. 1990. An analysis of curvature in fast and slow human saccades. *Experimental Brain Research* 81, 2 (1990), 335–345.
- Oleg Špakov and Yulia Gizatdinova. 2014. Real-time hidden gaze point correction. In *Proceedings of the symposium on eye tracking research and applications*. ACM, 291–294.
- 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*. ACM, 363–372.
- Yusuke Sugano, Yasuyuki Matsushita, and Yoichi Sato. 2013. Appearance-based gaze estimation using visual saliency. *IEEE transactions on pattern analysis and machine intelligence* 35, 2 (2013), 329–341.
- Yusuke Sugano, Yasuyuki Matsushita, Yoichi Sato, and Hideki Koike. 2015. Appearance-based gaze estimation with online calibration from mouse operations. *IEEE Transactions on Human-Machine Systems* 45, 6 (2015), 750–760.
- Steven P Tipper, Louise A Howard, and Stephen R Jackson. 1997. Selective reaching to grasp: Evidence for distractor interference effects. *Visual cognition* 4, 1 (1997), 1–38.
- Subarna Tripathi and Brian Guenter. 2017. A statistical approach to continuous self-calibrating eye gaze tracking for head-mounted virtual reality systems. In *Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on*. IEEE, 862–870.
- Luke Tudge, Eugene McSorley, Stephan A Brandt, and Torsten Schubert. 2017. Setting things straight: A comparison of measures of saccade trajectory deviation. *Behavior research methods* 49, 6 (2017), 2127–2145.
- Stefan Van der Stigchel. 2010. Recent advances in the study of saccade trajectory deviations. *Vision research* 50, 17 (2010), 1619–1627.
- Stefan Van der Stigchel, Martijn Meeter, and Jan Theeuwes. 2006. Eye movement trajectories and what they tell us. *Neuroscience & biobehavioral reviews* 30, 5 (2006), 666–679.
- AJ Van Opstal and JAM Van Gisbergen. 1987. Skewness of saccadic velocity profiles: a unifying parameter for normal and slow saccades. *Vision research* 27, 5 (1987), 731–745.

- Paolo Viviani, Alain Berthoz, and David Tracey. 1977. The curvature of oblique saccades. *Vision research* (1977).
- Robin Walker, Eugene McSorley, and Patrick Haggard. 2006. The control of saccade trajectories: Direction of curvature depends on prior knowledge of target location and saccade latency. *Perception & Psychophysics* 68, 1 (2006), 129–138.
- Kang Wang, Shen Wang, and Qiang Ji. 2016. Deep eye fixation map learning for calibration-free eye gaze tracking. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*. ACM, 47–55.
- Zhiguo Wang, Jason Satel, Thomas P Trappenberg, and Raymond M Klein. 2011. Aftereffects of saccades explored in a dynamic neural field model of the superior colliculus. *Journal of Eye Movement Research* 4, 2 (2011).
- Juan Xu, Ming Jiang, Shuo Wang, Mohan S Kankanhalli, and Qi Zhao. 2014. Predicting human gaze beyond pixels. *Journal of vision* 14, 1 (2014), 28–28.
- AL Yarbus. 1967. Eye movements and vision. 1967. *New York* (1967).
- Xucong Zhang, Michael Xuelin Huang, Yusuke Sugano, and Andreas Bulling. 2018. Training Person-Specific Gaze Estimators from User Interactions with Multiple Devices. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 624.