Prediction of Gaze Estimation Error for Error-Aware Gaze-Based Interfaces

Michael Barz[1] Florian Daiber[1,2] Andreas Bulling[3]
[1]German Research Center for Artificial Intelligence (DFKI), Germany {michael.barz,florian.daiber}@dfki.de
[2]Lancaster University, United Kingdom
[3]Perceptual User Interfaces Group, Max Planck Institute for Informatics, Germany bulling@mpi-inf.mpg.de

Abstract

Gaze estimation error is inherent in head-mounted eye trackers and seriously impacts performance, usability, and user experience of gaze-based interfaces. Particularly in mobile settings, this error varies constantly as users move in front and look at different parts of a display. We envision a new class of gaze-based interfaces that are aware of the gaze estimation error and adapt to it in real time. As a first step towards this vision we introduce an error model that is able to predict the gaze estimation error. Our method covers major building blocks of mobile gaze estimation, specifically mapping of pupil positions to scene camera coordinates, marker-based display detection, and mapping of gaze from scene camera to on-screen coordinates. We develop our model through a series of principled measurements of a state-of-the-art head-mounted eye tracker.

CR Categories: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Evaluation/methodology

Keywords: Eye Tracking; Gaze Estimation; Error Modelling and Prediction; Gaze-Based Interaction; Error-Aware Interfaces

1 Introduction

Recent advances in head-mounted eye tracking promise gaze-based interaction with ambient displays in pervasive daily-life settings [Bulling and Gellersen 2010]. A key problem in mobile settings is that gaze estimation error, i.e. the difference between the estimated on-screen and the true gaze position, is often substantial while the user moves in front of one or multiple displays [Lander et al. 2015]. Several methods were proposed to address this problem, such as filtering gaze jitter or snapping gaze to on-screen objects [Špakov 2012; Špakov and Gizatdinova 2014]. More recent works try to reduce gaze estimation error, e.g. through continuous self-calibration [Sugano and Bulling 2015]. Although they can improve user experience, all of these approaches only alleviate the symptoms and do not aim to embrace the inevitable gaze estimation error in the interaction design. These approaches also don't allow designers of interactive systems to simulate gaze estimation error or to predict the error to adapt pro-actively during runtime and depending on the current user position, orientation and on-screen gaze position. As a consequence, current interfaces do not leverage the full potential of gaze input.

ETRA '16,, March 14 - 17, 2016, Charlston, SC, USA
ISBN: 978-1-4503-4125-7/16/03. . .$15.00
DOI: http://dx.doi.org/10.1145/2857491.2857493

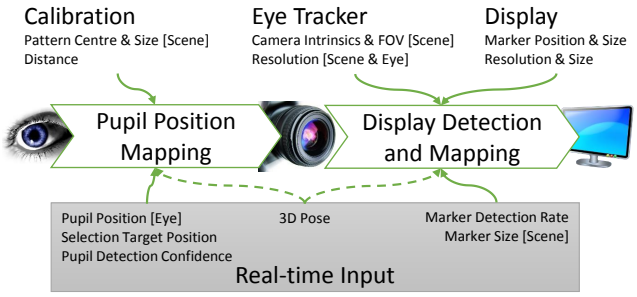

Figure 1: *Gaze estimation error model for head-mounted eye trackers comprising two main building blocks: Pupil Position Mapping and Display Detection and Gaze Mapping. Model inputs include parameters for Calibration, Eye Tracker and Display, as well as real-time gaze, visual marker, and 3D head pose, some specific to the Eye or Scene camera.*

We envision a novel class of mobile gaze-enabled interfaces that are "aware" of the gaze estimation error. This enables interfaces to adapt to gaze estimation error at runtime, for example, by magnifying on-screen objects in high-error regions or by moving them to low-error region of the display. As a key building block for these error-aware interfaces, in this work we present a method to model and predict gaze estimation error for head-mounted eye trackers depending on the current user position and orientation as well as on-screen gaze position. Our method models error of the major processing steps for mobile gaze estimation, specifically mapping of pupil positions to scene camera coordinates, marker-based display detection, as well mapping of gaze from scene camera to display coordinates (see Figure 1). Input to our model are 1) properties of the display, such as physical size and resolution, and the visual markers, 2) intrinsics of the eye tracker cameras, 3) parameters of the calibration routine and pattern, as well as 4) the user's current position and orientation relative to the display. The model provides real-time output of the gaze estimation error for any 3D surface in the field of view (FOV) of the eye tracker's scene camera specified by visual markers attached to the surface.

The specific contributions of this work are twofold. First, we report a series of measurements to characterise the error for building blocks commonly used in mobile gaze interaction using head-mounted eye trackers. Specifically, we quantify extrapolation and parallax error, error for detecting the display in the scene camera, and for mapping gaze coordinates from the scene camera to the display. Second, we present a support vector regression model that can predict gaze estimation error in real time depending on the current user position, orientation, and on-screen gaze position.

2 Modelling Gaze Estimation Error

Monocular head-mounted eye trackers are typically equipped with two cameras: a scene camera that captures part of the user's current FOV, and an eye camera that records a close-up video of the user's eye [Kassner et al. 2014]. The problem of gaze estimation is that

of mapping 2D pupil positions in the eye camera coordinate system to 2D gaze positions in the scene camera coordinate system [Majaranta and Bulling 2014]. The mapping is usually established in a calibration process. Pupil positions and corresponding scene camera positions are then typically mapped to each other using a first or second order polynomial. If these gaze positions are to be used for interacting with a display placed somewhere in the environment they have to be mapped further to the corresponding display coordinate system, e.g. by using visual markers attached to the display [Yu and Eizenman 2004] or by detecting the display itself [Mardanbegi and Hansen 2011]. This indicates two main components where errors can arise (see Figure 1): 1) the mapping of 2D pupil positions in eye camera coordinates to 2D scene camera coordinates (*Pupil Position Mapping*), as well as 2) detecting interactive displays in the environment and mapping gaze from scene camera coordinates to display coordinates (*Display Detection and Mapping*).

We focus on extrapolation and parallax error for *Pupil Position Mapping*. These errors are particularly important for mobile gaze interaction where users frequently change their position in front of a display [Cerrolaza et al. 2012; Mardanbegi and Hansen 2012]. We assume measures of the pupil detection error to be provided by the manufacturer, such as the pupil detection confidence c value provided by the PUPIL eye tracker [Kassner et al. 2014]. In addition, the detection of ambient displays is essential for gaze-based interaction and commonly combined with homographies to map the gaze estimates to that display [Breuninger et al. 2011]. Errors caused by the display detection algorithm propagate when mapping gaze and thus are covered by the *Display Detection and Mapping* block. We use separate models for the error in x and y direction as proposed by [Holmqvist et al. 2012]. The resulting model has several input parameters that are described in detail in [Barz et al. 2015]. S_p is the ratio between the calibrated and the total scene camera area. Normalised by S_p we define d_t^x (d_t^y) as difference between scene targets T_x (T_y) and calibration centre C_{cal}^x (C_{cal}^y). To model vergence we propose d_p^{rel}, the difference between recording and calibration distance, normalised by the squared calibration distance.

3 Pupil Position Mapping Error

We first studied extrapolation and parallax error independently to quantify their contribution to the pupil position mapping error.

Extrapolation Error. To quantify the extrapolation error, d_{cal} and d_{rec} were fixed to 250 cm while the size of the calibration pattern was varied between 100%, 75%, and 50% influencing S_p. We asked users to calibrate the eye tracker three times, each followed by one recording in which they looked at 13 target locations (T_x, T_y) equally distributed across the scene camera's FOV.

Parallax Error. To quantify the parallax error, we varied the distance between user and display during calibration d_{cal} and recording d_{rec}. The edge size of the calibration pattern was changed accordingly, i.e. in such a way that its relative size S_p remained constant. Similar to the first measurement users were asked to calibrate the system three times from certain positions $d_{cal} \in \{100$ cm, 200 cm, 250 cm$\}$. After each calibration users performed three recordings, one at the current calibration distance and two at the other distances, i.e. $|d_{rec} - d_{cal}| \in \{0, 50, 100, 150\}$ cm. The target locations were the same as for the first measurement.

3.1 Experimental Setup and Procedure

We recruited 12 participants (five female), aged between 19 and 50 years ($M = 24.067$, $SD = 7.459$), each receiving 15 EUR as compensation. Two participants for measurement 1 and one for measurement 2 had to be excluded from the analysis due to problems with the eye tracker. To record gaze we used a PUPIL head-mounted eye tracker [Kassner et al. 2014]. The monocular device features a scene camera with a resolution of 720p and an eye camera with 640×480 pixels, both delivering videos at 30 fps. The scene camera has a FOV of 90 degrees. To show the stimuli we used a projector mounted at the ceiling with a resolution of 1400×1050 pixels with a corresponding size on the canvas of 267×200 cm.

Participants were first introduced to the experiment and asked to complete a questionnaire on demographics and prior eye tracking experience. Their heads were then fixed using a chin rest. For both measurements, participants first calibrated the eye tracker using PUPIL's standard on-screen calibration. Afterwards, they were asked to look at a cross-hair that was manually positioned by the experimenter indicating the 13 stimuli one after another. The recording took on average 50 minutes per participant.

	M		SD	
100%	28.19 px	1.92°	7.67 px	0.55°
75%	28.07 px	1.87°	6.77 px	0.46°
50%	28.34 px	1.87°	6.29 px	0.39°
0 cm	24.97 px	1.68°	4.86 px	0.34°
±50 cm	27.76 px	1.9°	5.54 px	0.38°
±100 cm	31.5 px	2.13°	3.89 px	0.26°
±150 cm	35.87 px	2.44°	6.86 px	0.48°

Table 1: *Mean and standard deviation of spatial accuracy in scene camera coordinates for measurements on extrapolation error (top) and parallax error (bottom)*

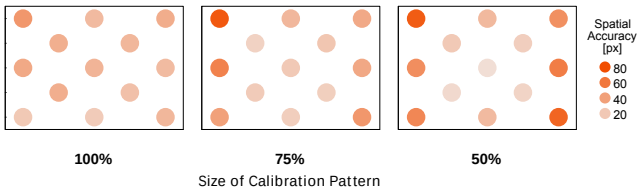

Figure 2: *Spatial accuracy in scene camera coordinates for all targets and different sizes of the calibration pattern.*

3.2 Results

Extrapolation Error. We first analysed the spatial accuracy for each sub-condition of the first measurement averaged over all scene targets (see Table 1 top). A repeated measures ANOVA ($N = 10$) showed no significant difference for the corresponding means ($F(2,8) = .007, p = .993$). We therefore analysed the gaze estimation error separately for each scene target location. As can be seen from Figure 2, the spatial accuracy for the full-sized pattern was evenly distributed across the display. For patterns with 75% and 50% edge length, the error at the display borders (8 outer stimuli) increased by 33.15% and 56.18%, respectively, while it decreased in the centre by 37.27% and 51.02% compared to the full pattern. A further ANOVA test showed a significant difference for the display border ($F(2,8) = 7.832, p = .013$) and the display centre ($F(2,8) = 15.715, p = .002$). A pairwise comparison (bonferroni-corrected) showed that the means of condition 100% are significantly different to 75% and 50%, whereas means of condition 75% are not significantly different when compared to 50% for both, stimuli at border and centre.

Parallax Error. We first grouped the data with respect to the absolute difference in calibration and recording distance $|d_{rec} - d_{cal}|$, ranging from 0 cm to 150 cm (50 cm increments), see Table 1 bottom. For spatial accuracy, an ANOVA test ($N = 11$) showed that these differences are significant ($F(3, 8) = 9.041, p = 0.006$). Accordingly a movement of 50 cm after calibration results in a decrease of spatial accuracy of 11.17% (26.15% for 100 cm, 43.65% for 150 cm). A pairwise comparison for 0 cm showed that all differences in means are significant. Apart from that only the differences between 50 cm and 150 cm are significant.

4 Display Detection and Mapping Error

Marker-based display detection and tracking to calculate a mapping from scene camera coordinates to display coordinates is increasingly used for gaze-based interaction (e.g. [Yu and Eizenman 2004; Breuninger et al. 2011]). Still, existing works typically considered marker detection and tracking as a black box system and did not quantify its contribution to gaze estimation error. While the specific error contribution, of course, depends on the particular markers and tracking algorithm used, it remains interesting to study one sample system and its interplay with other parts of the gaze estimation pipeline. The distance and orientation between scene camera and display, as well as the number and size of the markers, are important for robust marker detection and therefore potential sources of error. To complement the extrapolation and parallax error measurements, we performed another measurement on the error stemming from the display detection and gaze mapping to that display.

4.1 Experimental Setup and Procedure

Because gaze mapping is independent of the gaze estimation in scene camera coordinates, recording a dataset of scene images from different positions in front of the display without participants was sufficient. We recorded these images using the PUPIL head-mounted eye tracker and a wall-mounted 50-inch flat screen with a resolution of 1920×1080 pixels (17.42 px/cm). The eye tracker was mounted on a tripod and precisely positioned at predefined locations using an attached plumb-line. The ArUco library [Garrido-Jurado et al. 2014] was used for marker detection and tracking.

To record the dataset, we systematically varied the distance $d_{rec} \in \{75 \text{ cm}, 100 \text{ cm}, 200 \text{ cm}, 300 \text{ cm}\}$ and orientation α (pitch) and β (yaw) $\in \{0°, 20°, 40°, 60°\}$ of the eye tracker to the display. The roll angle was assumed to be zero. We recorded 800 images for all 84 combinations, totalling 67200 samples. In a post-hoc analysis we manually annotated all images with the corresponding display centre position in scene camera coordinate space. One image per physical location was used as a reference. The display centres in these reference images were mapped to display space using the homography matrix obtained during the recording session. Under ideal conditions, these points should be mapped to the centre of the display which was used as ground-truth.

4.2 Results

Figure 3a shows the error for mapping 2D gaze positions in scene camera coordinates to display coordinates for different angles and distances to the display. The mapping error increases with increasing angle and distance. Figure 3b plots the mean gaze estimation error and the marker detection rate against the distance. The smallest absolute error of 5.369 px ($SD = 7.277$) [$M = 0.31$ cm ($SD = 0.42$); $M = 0.18°$ ($SD = 0.24$)] was achieved at a distance of 100 cm with a detection rate of nearly 100%. The smallest error in degrees of visual angle was achieved at 200 cm with 0.11° ($SD = 0.06°$) [$M = 6.51$ px ($SD = 3.42$); $M = 0.37$ cm

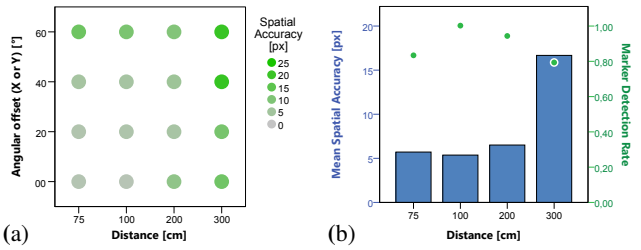

Figure 3: *Gaze estimation error for the display mapping for different angles and distances to the display in display coordinate space (a). Relations between the distance and error and between the distance and the marker detection rate (b).*

($SD = 0.2$)]. For distances beyond 100 cm, detection rate decreases and the absolute error increases. Visual inspection of the scene videos showed that the small increase at 75 cm was due to not all markers having been visible in the scene camera's FOV.

5 Evaluation of the Combined Error Model

The previous experiments provided important insights into how much extrapolation and parallax error as well as display detection and mapping individually contribute to the overall gaze estimation error. We performed an evaluation to assess the performance of the combined error model, i.e. the model combining the *Pupil Position Mapping* and the *Display Detection and Mapping* components. We trained two support vector regression (SVR) models with radial basis function (RBF) kernels on the datasets recorded during experiments 1 and 2 – one model for pupil mapping error and one for display mapping error. The data were partitioned into a training and a test set each (70%/30%) and used to evaluate the model.

For evaluation, the model first estimated the error caused by the *Pupil Position Mapping* component. The result was an error estimate in scene camera space that we transferred to display coordinates with a distance dependent mapping. Afterwards the display mapping error was predicted in display coordinates and added one-to-one to the prior result. Model performance is reported by means of root mean squared error (RMSE) of the prediction residuals and R^2 as a measure for the portion of the data variance explained by the given model. This test was repeated on 50 randomly chosen training/test sets (see [Barz et al. 2015] for details).

The performance of the combined error model – as a function of the distance to the display – is shown in Figure 4. On average, the model achieved an overall spatial accuracy from 3.96 px [0.75 cm] at 50 cm to 23.74 px [4.53 cm] at 300 cm in x direction, and 2.36 px [0.45 cm] at 50 cm to 14.19 px [2.71 cm] at 300 cm in y direction. These values correspond to 0.86° for the x-model and to 0.52° for the y-model. In addition, we compared our model to two baseline approaches. *Best* assumes a constant error of 0.6°, which is reported as best-case spatial accuracy of the PUPIL tracker [Kassner et al. 2014]. The *Measured* model takes the mean error in visual degrees extracted from our measurements as a basis. The means are 1.26° for both x and y direction. To simulate the residuals for *Best* and *Measured* we calculated gaze error estimates dependent on the distance and compared them to the same test set as used for evaluating our model.

6 Discussion

As a key building block for a novel class of error-aware interfaces, in this work we presented a method to model and predict the gaze

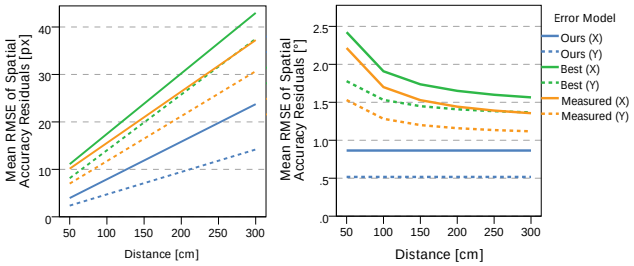

Figure 4: *Error estimation performance in cm and degrees of visual angle in x and y direction of the proposed error model (Ours), best-case model (Best), and measured model (Measured) for different distances to the display.*

estimation error of head-mounted eye trackers in real time. Results from our study suggest that the chosen set of inputs is comprehensive and allows the model to predict the gaze estimation error with a RMSE of 0.86° for x and 0.52° for y. To the best of our knowledge, this is the first attempt to develop such a model, characterise its inputs and evaluate its performance.

Although calibration pattern size, distance as well as display detection and mapping are known sources of error for mobile gaze interaction, this work is also first to quantify how each of these sources contribute to overall gaze estimation error. Specifically, our evaluations revealed that in mobile settings extrapolation error is significant and that a denser calibration pattern can result in better accuracy. This finding is particularly important for gaze-based interfaces as it demonstrates that our model could, e.g., be used to create high-accuracy regions for fine-grained interactions traded off with lower accuracy in other regions. Our second measurement extends on previous findings in that we not only confirm that parallax error is a significant source of error but also to which extend.

Despite its advantages in terms of performance and usability our model also has limitations that need to be studied in future work. First, we currently train two separate models for error in x and y direction. While this approach was shown to work well [Holmqvist et al. 2012], we believe that a single model that outputs a joint error for both directions would be preferable. Second, future work could study additional error sources, such as displacement of the eye tracker on the head that was shown to be important, particularly for long-term recordings in mobile settings, or motion blur caused by fast head movements, which can impact marker detection performance. Third, binocular eye tracking may decrease the parallax error and will therefore be interesting to investigate and incorporate in a future extension of our model.

7 Conclusion

We proposed a novel method to model and predict the error inherent for head-mounted eye trackers. This enables a new class of gaze-based interfaces that are aware of the gaze estimation error and driven by real-time error estimation. We performed a series of measurements that provided important insights into the individual error contribution of major building blocks for mobile gaze estimation. Results from our study suggest that the chosen set of inputs is comprehensive and allows to predict the gaze estimation error with a reasonable accuracy.

Acknowledgements

This work was funded, in part, by the Cluster of Excellence on Multimodal Computing and Interaction (MMCI) at Saarland University, the Deutsche Forschungsgemeinschaft (DFG KR3319/8-1) and the EU Marie Curie Network iCareNet under grant agreement 264738.

References

BARZ, M., BULLING, A., AND DAIBER, F. 2015. Computational modelling and prediction of gaze estimation error for head-mounted eye trackers. Tech. rep., German Research Center for Artificial Intelligence (DFKI).

BREUNINGER, J., LANGE, C., AND BENGLER, K. 2011. Implementing gaze control for peripheral devices. In *Proc. PETMEI*, 3–8.

BULLING, A., AND GELLERSEN, H. 2010. Toward mobile eye-based human-computer interaction. *IEEE Pervasive Computing 9*, 4, 8–12.

CERROLAZA, J. J., VILLANUEVA, A., VILLANUEVA, M., AND CABEZA, R. 2012. Error characterization and compensation in eye tracking systems. In *Proc. ETRA*, 205–208.

GARRIDO-JURADO, S., NOZ SALINAS, R. M., MADRID-CUEVAS, F., AND MARÍN-JIMÉNEZ, M. 2014. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition 47*, 6, 2280 – 2292.

HOLMQVIST, K., NYSTRÖM, M., AND MULVEY, F. 2012. Eye tracker data quality: What it is and how to measure it. In *Proc. ETRA*, 45–52.

KASSNER, M., PATERA, W., AND BULLING, A. 2014. Pupil: An open source platform for pervasive eye tracking and mobilegaze-based interaction. In *Adj. Proc. UbiComp*, 1151–1160.

LANDER, C., GEHRING, S., KRÜGER, A., BORING, S., AND BULLING, A. 2015. Gazeprojector: Accurate gaze estimation and seamless gaze interaction across multiple displays. In *Proc. UIST*, 395–404.

MAJARANTA, P., AND BULLING, A. 2014. *Eye Tracking and Eye-Based Human-Computer Interaction*. Advances in Physiological Computing. Springer, 39–65.

MARDANBEGI, D., AND HANSEN, D. W. 2011. Mobile gaze-based screen interaction in 3d environments. In *Proc. NGCA*, 2.

MARDANBEGI, D., AND HANSEN, D. W. 2012. Parallax error in the monocular head-mounted eye trackers. In *Proc. UbiComp*, 689–694.

ŠPAKOV, O., AND GIZATDINOVA, Y. 2014. Real-time hidden gaze point correction. In *Proc. ETRA*, 291–294.

ŠPAKOV, O. 2012. Comparison of eye movement filters used in hci. In *Proc. ETRA*, 281–284.

SUGANO, Y., AND BULLING, A. 2015. Self-calibrating head-mounted eye trackers using egocentric visual saliency. In *Proc. UIST*, 363–372.

YU, L. H., AND EIZENMAN, E. 2004. A new methodology for determining point-of-gaze in head-mounted eye tracking systems. *IEEE Transactions on Biomedical Engineering 51*, 10, 1765–1773.