Solving Parallax Error for 3D Eye Tracking

Agostino Gibaldi
agostino.gibaldi@berkeley.edu
UC Berkeley School of Optometry
and Vision Science Program
Berkeley, CA, USA

Vasha DuTell
vasha@berkeley.edu
UC Berkeley School of Optometry and
Vision Science Program and Redwood
Center for Theoretical Neuroscience
Berkeley, CA, USA

Martin S. Banks
martybanks@berkeley.edu
UC Berkeley School of Optometry
and Vision Science Program
Berkeley, CA, USA

ABSTRACT

Head-mounted eye-trackers allow for unrestricted behavior in the natural environment, but have calibration issues that compromise accuracy and usability. A well-known problem arises from the fact that gaze measurements suffer from parallax error due to the offset between the scene camera origin and eye position. To compensate for this error two pieces of data are required: the pose of the scene camera in head coordinates, and the three-dimensional coordinates of the fixation point in head coordinates. We implemented a method that allows for effective and accurate eye-tracking in the three-dimensional environment. Our approach consists of a calibration procedure that allows to contextually calibrate the eye-tracker and compute the eyes pose in the reference frame of the scene camera, and a custom stereoscopic scene camera that provides the three-dimensional coordinates of the fixation point. The resulting gaze data are free from parallax error, allowing accurate and effective use of the eye-tracker in the natural environment.

CCS CONCEPTS

• **Human-centered computing** → **Interaction devices**.

KEYWORDS

wearable eye tracking, eye tracking calibration, parallax error

ACM Reference Format:

Agostino Gibaldi, Vasha DuTell, and Martin S. Banks. 2021. Solving Parallax Error for 3D Eye Tracking. In *2021 Symposium on Eye Tracking Research and Applications (ETRA '21 Adjunct)*, May 25–27, 2021, Virtual Event, Germany. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3450341.3458494>

1 INTRODUCTION

A standard approach to head-mounted eye-tracking consists of defining a transformation between pupil position in eye-tracking camera coordinates and a pixel position of the scene camera image. This transformation is estimated with a calibration procedure that varies, depending on the accuracy required, from a one-point calibration to three, five, nine, or even thirteen points [Gibaldi et al. 2017b; Holmqvist et al. 2011].

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 the author(s) 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.

ETRA '21 Adjunct, May 25–27, 2021, Virtual Event, Germany

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-8357-8/21/05...\$15.00

<https://doi.org/10.1145/3450341.3458494>

Despite its simplicity, this approach requires the relation between the head position and the the visual scene to remain unchanged. If the scene - specifically the three-dimensional (3D) structure of the scene - changes, the transformation will suffer from parallax error [Duchowski et al. 2014; Elmadjian et al. 2018; Kothari et al. 2020; Mardanbegi and Hansen 2012] caused by the offset between the scene camera origin and eye position (see Fig. 1).

In order to properly perform eye-tracking in the natural, unstructured environment, two conditions are necessary: first, it is necessary to know the transformation between the scene camera and the eye system; second, it is necessary to reconstruct the 3D structure of the visual scene. On the basis of these, it is possible to perform a correct and accurate 3D estimation of gaze direction for a head mounted eye-tracker.

Other researchers mitigated the problem by estimating a 3D model of the eyes at calibration [Kassner et al. 2014; Mansouryar et al. 2016], even if the 3D scene structure is not measured. An outstanding example that resorts on a Depth camera and performs the calibration on the fly is provided in [Kothari et al. 2020]. Calibration is performed using a checkerboard grid to contextually compute the transformation between the RGB and depth cameras and the eye-tracker calibration. Despite the effectiveness and usability of the approach, it is worth noting that the method does not estimate the eye position, thus only mitigating the parallax problem.

Here we present a practical implementation to contextually measure eye position and 3D scene structure, for 3D eye-tracking in the natural environment, which has been successfully used in various experiments [Gibaldi and Banks 2019, 2021b; Gibaldi et al. 2017a, 2021; Sprague et al. 2015].

2 METHODS AND RESULTS

Related Works. We built a custom scene- and eye-tracking apparatus to measure both the 3D structure of the scene as well as subject's fixation point in those scenes as they engaged in everyday activities.

Apparatus. The apparatus consisted of head-mounted stereoscopic camera pair, used as a scene camera, as well as a binocular eye tracker to measure gaze direction for each eye. The eye tracker was the SR Research EyeLink II, configured to measure gaze direction using pupil position and corneal reflections, running at 250Hz. The infra-red head camera, generally required for computing the screen position, was not used, as we implemented a fixed setup with accurate, known positions. The cameras were two Sony XCD-MV6 digital video cameras recording in gray scale at 30Hz with a resolution of 640×480 . The field of view was 75×58 degrees.

Figure 1: Parallax error in head-mounted eye-tracking. Panel A shows the side view of a head-mounted eye-tracker. When the scene object is at a similar distance to the calibration distance, the device provides an accurate estimation of gaze direction. When the scene object is at a different distance, the parallax error between the scene camera and the eyes results in a systematic bias of the gaze estimation. This bias increases with distance between the scene object and calibration position, both in front and behind. Panel B shows a zoom of the side view, together with the orthonormal bases representing the camera pose e^c and the eye pose e^e , and the transformation τ required to transform e^c into e^e .

The two cameras were synchronized in a master/slave configuration, and the trigger signal was a hardware signal generated by the EyeLink II and sent by the cameras by a dedicated serial port. The intrinsic and extrinsic parameters of the stereo rig were computed using standard approaches available in OpenCV [Zhang 2000]. More details about the apparatus can be found in [Sprague et al. 2015].

Eye-Tracking Calibration Setup and Procedure. Each session began with the standard calibration provided by SR Research. Subjects sat in front of a large front-to-parallel LCD screen (125×77 cm) with HD resolution (1920×1080 pixels). The subject's head was stabilized with a bite bar at 100 cm from the screen. The eyes were positioned very accurately relative to the screen using a sighting device [Hillis and Banks 2001]. The midpoint of the subject's inter-ocular axis intersected a surface normal from the center of the display screen (see Fig. 2A).

Thirteen calibration positions were presented followed by nine new positions for validation [Gibaldi et al. 2017b]. In such a configuration, the calibrated range was at the working limit of EyeLink II, that is at $\approx \pm 30$ degrees horizontally and $\approx \pm 20$ degrees vertically. To ensure accuracy, calibration was repeated until the validation yielded a mean error less than 0.5° . Since the SR Research software allows for separate monocular calibrations, each eye was calibrated separately in order to improve accuracy for binocular tasks [Gibaldi et al. 2017b; Švede et al. 2015].

The eye-tracker head-band was positioned such that the calibration screen was centered in the stereo scene camera images. The transformation τ between the stereo scene camera and the eye system (see Fig. 1B), is made of two components, a translation t to move the scene camera center O^c to the eye center O^e , and a rotation R , to align the orthonormal basis e^c of the scene camera to that of the eye e^e . In order to compute it, a grid pattern was displayed on the screen immediately after the eye-tracking calibration (see Fig. 2B). Standard approaches available in OpenCV [Zhang 2000] were then used to compute the $\tau = [R|t]$.

This procedure allowed for reporting both the gaze data for each eye and the 3D scene structure within the same reference frame. For

convenience, we set this at the level of the cyclopean eye, oriented towards the screen center.

3D Gaze Estimation. After the calibration procedure, the subject was free to step off the bite-bar and to move in the natural environment. The stereo images were used to generate a depth map of the visual scene using an established stereo reconstruction algorithm [Hirschmüller and Scharstein 2008]. The intrinsic and extrinsic parameters of the stereo rig were used to compute a 3D point cloud of the scene. Then the estimated relative pose of the eye system in camera coordinates was used to place the binocular fixation point in the 3D scene (see Fig. 2C).

The eye-tracker returned the fixation points for the left and right eyes in screen pixel coordinates (green and red crosses). Even if the screen is not physically present, knowing the physical size, distance and orientation of the screen allows the transform of these coordinates in 3D space. The left and right fixation axes (green and red dashed lines) were computed as the 3D lines connecting these screen points with each eye. The binocular fixation point was the best intersection of the two axes, and the cyclopean fixation axis (yellow dashed line) was the 3D line connecting this point and the cyclopean eye. Finally, the 3D position of cyclopean fixation point was refined using the 3D structure of the scene, as the point along the cyclopean fixation axis intersecting the point in the scene closest to the subject (yellow cross on cup).

Validation. In order to evaluate the accuracy and precision of the proposed approach, we implemented a validation procedure. The procedure is performed once after the eye-tracker calibration and before performing the experiment, in order to check if it provided sufficient accuracy, and once after the experiment, to control for possible slippage of the device. The procedure was similar to a standard validation performed for eye-tracking calibration: a total of XX arranged on a polar grid (eccentricities of X, Y and Z degrees, and 6 orientation) were showed in random sequence on screen. The subject, still on the bite-bar, was instructed to gaze the displayed target and move to the next with a button press. Thanks to the accurate positioning of the subject with respect to the screen, the actual gaze angles (azimuth, elevation and vergence) were known

Figure 2: Solution of parallax error in head-mounted eye-tracking. Panel A. Calibration setup consisting of the head-mounted device (binocular eye-tracker and stereoscopic scene camera), a bite-bar stand, and a large LCD screen at a known distance and orientation. Accurate positioning is ensured with a custom gun-sight device [Hillis and Banks 2001]. On the screen is shown the position of the fixation targets used for a thirteen-point calibration. The green and red dashed lines represent the fixation axes for the left and right eye, respectively. The yellow dot on the head represents the position of the cyclopean eye, and the dashed yellow line is its fixation axis. **Panel B. Measurement of the relative pose** between the stereo scene camera and the eye system. On the screen is shown a camera calibration grid while the bite-bar allows the subject to assume the same position used during the eye-tracking calibration. **Panel C. Usage of the device.** The subject is no longer on the bite-bar, and moves freely in the natural environment. Data from the stereo scene camera and the eye-tracker are used to reconstruct the 3D structure of the scene and to place the binocular fixation point in it (yellow cross on cup).

for fixation each target and compared with the those measured with the proposed approach. If either validations provided a root mean square error > 0.8 degrees, the experimental session was voided.

Results. The approach has been successfully used to create the BinOcular Image Statistics DATABASE (BORIS) dataset, a collection of gaze and retinal disparity for fixating observers performing everyday tasks in the natural environment [Gibaldi and Banks 2021a]. Four subjects performed seven tasks each, from making a sandwich to walking around Campus. The average accuracy of gaze data was measured to be 0.71 ± 0.14 degrees (mean $\pm$ standard deviation). About one third of the trials were discarded due to high error in the validation procedure performed after the trial.

The data from the BORIS dataset have been proven useful to interpret functional properties of the visual system, specifically related to stereopsis [Gibaldi and Banks 2021b; Sprague et al. 2015], binocular coordination [Gibaldi and Banks 2019] and the optical properties of the eye [Gibaldi et al. 2021].

3 CONCLUSION AND FUTURE DIRECTIONS

In this work, we describe the guidelines for a method of 3D gaze estimation using head-mounted eye-trackers in natural conditions. Gaze data as pixel positions on the (virtual) calibration screen stay in a fixed reference frame with respect to the head. Measuring the relative position between the screen, the eyes and the scene camera makes it possible to transform these data into 3D coordinates, expressed in the same head-centered reference frame. Measuring the 3D structure of the scene allows one to mitigate eye-tracking imprecision, allowing reliable measurements of eye vergence.

Despite its effectiveness, the presented approach has some drawbacks. Most notably, the calibration is long, somewhat fatiguing, and must be performed in a specially-designed laboratory setup, which limits the usability of this method in some settings.

Future Directions. Considering the recent widespread adoption of head-mounted eye-tracking and depth cameras, we aim at implementing this approach in a simple and effective solution for calibration of such systems with head-free movements. To improve usability, the approach has to be performed outside a lab setting, directly on the site of the experiment (e.g. see [Kothari et al. 2020]). Such approach would provide accurate 3D gaze data in all studies requiring changeable and unpredictable fixation distances.

ACKNOWLEDGMENTS

This work was supported by the Center for Innovation in Vision and Optics and the National Defense Science and Engineering Graduate Fellowship.

REFERENCES

- Andrew T Duchowski, Donald H House, Jordan Gestring, Robert Congdon, Lech Świrski, Neil A Dodgson, Krzysztof Krejtz, and Izabela Krejtz. 2014. Comparing estimated gaze depth in virtual and physical environments. In *Proceedings of the Symposium on Eye Tracking Research and Applications*. 103–110.
- Carlos Elmadjian, Pushkar Shukla, Antonio Diaz Tula, and Carlos H Morimoto. 2018. 3D gaze estimation in the scene volume with a head-mounted eye tracker. In *Proceedings of the Workshop on Communication by Gaze Interaction*. 1–9.
- Agostino Gibaldi and Marty Banks. 2021a. BinOcular Image Statistics (BORISa) Database. osf.io/t9qg5
- Agostino Gibaldi and Martin S Banks. 2019. Binocular eye movements are adapted to the natural environment. *Journal of Neuroscience* 39, 15 (2019), 2877–2888.
- Agostino Gibaldi and Martin S Banks. 2021b. Crossed–uncrossed projections from primate retina are adapted to disparities of natural scenes. *Proceedings of the National Academy of Sciences* 118, 7 (2021), e2015651118.
- Agostino Gibaldi, Andrea Canessa, and Silvio P Sabatini. 2017a. The Active Side of Stereopsis: Fixation Strategy and Adaptation to Natural Environments. *Scientific Reports* 7 (2017).
- Agostino Gibaldi, Vivek Labhishetty, Larry N Thibos, and Martin S Banks. 2021. The blur horopter: Retinal conjugate surface in binocular viewing. *Journal of Vision* 21, 8 (2021), –.
- Agostino Gibaldi, Mauricio Vanegas, Peter J Bex, and Guido Maiello. 2017b. Evaluation of the Tobii EyeX Eye tracking controller and Matlab toolkit for research. *Behavior research methods* 49, 3 (2017), 923–946.
- James M Hillis and Martin S Banks. 2001. Are corresponding points fixed? *Vision research* 41, 19 (2001), 2457–2473.
- Heiko Hirschmuller and Daniel Scharstein. 2008. Evaluation of stereo matching costs on images with radiometric differences. *IEEE transactions on pattern analysis and*

- machine intelligence* 31, 9 (2008), 1582–1599.
- 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.
- Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: Adjunct publication*. 1151–1160.
- Rakshit Kothari, Zhizhuo Yang, Christopher Kanan, Reynold Bailey, Jeff B Pelz, and Gabriel J Diaz. 2020. Gaze-in-wild: A dataset for studying eye and head coordination in everyday activities. *Scientific reports* 10, 1 (2020), 1–18.
- Mohsen Mansouryar, Julian Steil, Yusuke Sugano, and Andreas Bulling. 2016. 3D gaze estimation from 2D pupil positions on monocular head-mounted eye trackers. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*. 197–200.
- Diako Mardanbegi and Dan Witzner Hansen. 2012. Parallax error in the monocular head-mounted eye trackers. In *Proceedings of the 2012 acm conference on ubiquitous computing*. 689–694.
- William W Sprague, Emily A Cooper, Ivana Tošić, and Martin S Banks. 2015. Stereopsis is adaptive for the natural environment. *Science advances* 1, 4 (2015), e1400254.
- Aiga Švede, Elina Treija, Wolfgang Jaschinski, and Gunta Krūmiņa. 2015. Monocular versus binocular calibrations in evaluating fixation disparity with a video-based eye-tracker. *Perception* 44, 8-9 (2015), 1110–1128.
- Zhengyou Zhang. 2000. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence* 22, 11 (2000), 1330–1334.