A novel approach to single camera, glint-free 3D eye model fitting including corneal refraction

Kai Dierkes
Pupil Labs Research, Berlin
kd@pupil-labs.com

Moritz Kassner
Pupil Labs Research, Berlin
moritz@pupil-labs.com

Andreas Bulling
Pupil Labs Research, Berlin
andreas@pupil-labs.com

ABSTRACT

Model-based methods for glint-free gaze estimation typically infer eye pose using pupil contours extracted from eye images. Existing methods, however, either ignore or require complex hardware setups to deal with refraction effects occurring at the corneal interfaces. In this work we provide a detailed analysis of the effects of refraction in glint-free gaze estimation using a single near-eye camera, based on the method presented by [Świrski and Dodgson 2013]. We demonstrate systematic deviations in inferred eyeball positions and gaze directions with respect to synthetic ground-truth data and show that ignoring corneal refraction can result in angular errors of several degrees. Furthermore, we quantify gaze direction dependent errors in pupil radius estimates. We propose a novel approach to account for corneal refraction in 3D eye model fitting and by analyzing synthetic and real images show that our new method successfully captures refraction effects and helps to overcome the shortcomings of the state of the art approach.

CCS CONCEPTS

• **Computing methodologies** → **Computer vision**; **Model development and analysis**; *Tracking*;

KEYWORDS

Eye tracking, refraction, 3D eye model, pupil detection, contour-based, glint-free

ACM Reference Format:
Kai Dierkes, Moritz Kassner, and Andreas Bulling. 2018. A novel approach to single camera, glint-free 3D eye model fitting including corneal refraction. In *ETRA '18: 2018 Symposium on Eye Tracking Research and Applications, June 14–17, 2018, Warsaw, Poland.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3204493.3204525

1 INTRODUCTION

Over the last decades a wide variety of camera-based eye trackers have been proposed. Typically, these require active generation of glints by means of infrared (IR) LEDs and/or a calibrated pair of stereo cameras. However, in many use cases such as head-mounted eye trackers such a highly controlled and calibrated setup is hard or even impossible to achieve. In addition, one of the significant potentials of head-mounted eye tracking is its use in outdoor environments that pose challenges for robust generation of IR glints. From an application point of view it is thus highly desirable to design accurate and robust glint-free eye-tracking solutions based on a single camera only. While learning-based head-mounted eye trackers that only require off-the-shelf cameras have been proposed [Mayberry et al. 2014; Tonsen et al. 2017], such systems are still in their infancy. Recently, Świrski et al. presented an algorithmic framework to address the gaze estimation problem in this highly restrictive setting [Świrski and Dodgson 2013]. The authors developed a model-based approach, which at its core comprised the fit of a 3D eye model to a series of pupil contours extracted from images recorded with a single near-eye camera over time. Neglecting headset slippage occurring on long time scales [Sugano and Bulling 2015], such temporal integration of information is possible as for head-mounted eye trackers the point of rotation of the eyeball can be assumed to be fixed in the eye camera coordinate system.

A key challenge in contour-based gaze estimation is the refraction of light rays leaving or entering the eye through the cornea. Refraction is the bending of light rays at the interface of media with different refractive indices (see Figure 1A). The optical power of the eye stems to a large extent from the refractive properties of its various parts [Gross 2008]. As a consequence of refraction, the apparent shape of the pupil in an eye image is not a mere perspective projection of the actual 3D pupil, but the image of a virtual pupil (see Figure 1B), also referred to as the entrance pupil [Fedtke et al. 2010]. For remote systems that typically employ both

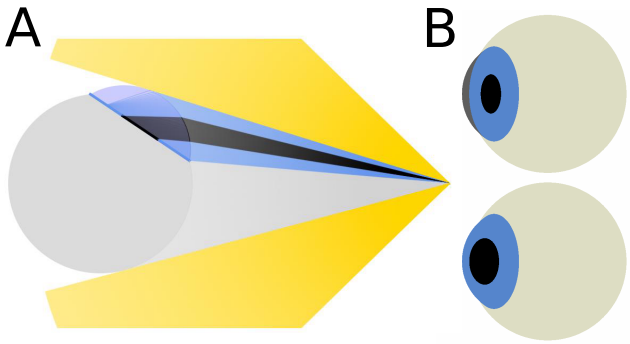

Figure 1: (A) Refraction leads to a bending of light rays entering or leaving the eye at the corneal interfaces. (B) Refraction effects pose a major challenge for glint-free gaze estimation methods given that apparent pupil contours depend on the refractive properties of the cornea. Top, no refraction. Bottom, spherical refractive cornea.

ETRA '18, June 14–17, 2018, Warsaw, Poland

ACM ISBN 978-1-4503-5706-7/18/06…$15.00
https://doi.org/10.1145/3204493.3204525

glint and pupil contour information, the effects of refraction have been noted and analyzed early on [Barsingerhorn et al. 2017; Villanueva and Cabeza 2008]. In particular, it has been shown that disregarding refraction can have a detrimental effect on angular gaze-estimation accuracy, with errors up to several degrees [Villanueva and Cabeza 2008]. For glint-free eye trackers such as the one proposed by [Świrski and Dodgson 2013], however, refraction effects have not been studied before.

The specific contributions of this work are two-fold. First, we provide a detailed assessment of the errors incurred by neglecting the influence of refraction within the framework developed by [Świrski and Dodgson 2013]. Employing a spherical 3D eye model that mathematically accounts for refraction occurring at corneal interfaces, we render corresponding synthetic eye images. Based on this pipeline we perform a simulation study, showing that disregarding refraction leads to an error in estimated (i) eyeball position, (ii) gaze vector, and (iii) pupil radius. Second, we introduce a novel glint-free approach to gaze estimation employing a single near-eye camera that accounts for corneal refraction by extending the algorithmic framework developed by [Świrski and Dodgson 2013]. To this end, we introduce a new cost function for (i) determining eyeball position in camera coordinates given a set of eye images by solving a nonlinear optimization problem, and (ii) for subsequently determining the gaze vector and pupil radius corresponding to individual eye images. We evaluate the performance of our approach by means of synthetic eye images and show that it correctly captures the effects of refraction. We furthermore analyze real eye images, verifying that the proposed optimization procedure is feasible and that the observed differences between the Świrksi model and our approach also pertain in a real world setting.

2 RELATED WORK

Our work is related to previous works on 1) model-based gaze estimation, specifically 2) contour-based approaches, as well as on 3) refraction effects in glint-based remote eye trackers.

2.1 Model-based gaze estimation

Video-based eye tracking can be broadly categorised depending on whether the method uses a regression or eye model approach [Hansen and Ji 2010]. Regression-based methods typically rely on polynomial mapping functions to predict 2D gaze coordinates on a screen or within the scene camera from eye image features [Fuhl et al. 2017, 2016; Javadi et al. 2015; Kassner et al. 2014; Mansouryar et al. 2016; Santini et al. 2018; Świrski et al. 2012; Tonsen et al. 2016]. Model-based approaches, in contrast, fit a 3D eye model to the eye images, allowing for gaze estimation in 3D space. In this work we focus on model-based gaze estimation methods.

Model-based methods exploit features extracted from eye and face images, recorded by one or more cameras, to determine pertinent parameters of the employed eye model. While there are gaze-estimation schemes which are based e.g. on facial features used for determining the position of the eyeball [Chen and Ji 2008], the majority of existing approaches relies on the analysis of one or more of the following features: pupil contours, pupil center positions, and glints. The latter are reflections on the corneal surface which are actively generated by means of one or more infrared (IR) LEDs [Hansen and Ji 2010]. The approach proposed in this work makes use of a single near-eye camera only, is based on pupil contours, and does not rely on glints.

2.2 Contour-based approaches

Tsukada et al. presented a 3D model-based gaze-estimation approach based on the fit of an eye model to iris contours extracted from eye images [Tsukada and Kanade 2012; Tsukada et al. 2011]. Their approach bears many similarities with the work presented in [Świrski and Dodgson 2013]. However, it remains unclear whether refraction effects undermine their basic premise, namely that iris contours are simple perspective projections of the circular 3D iris. While this appears to be reasonable for near-eye cameras mounted almost frontally to the eyes of a subject, for more oblique camera angles that are often required in head-mounted eye tracking because of headset geometry [Świrski et al. 2012], this assumption is not satisfied. We present work that does not make any assumptions of this kind. A contour based approach dealing with refraction in a non-approximative way was developed by Lai et al. but their method requires two cameras [Lai et al. 2015]. In this work, we are concerned with refraction modelling for gaze estimation using eye images from a single camera only.

2.3 Modelling refraction in glint-based remote eye trackers

Ohno et al. proposed a remote eye-tracking system consisting of a single camera and a single LED [Ohno et al. 2002]. Using a glint-based technique for determining the center of the cornea, they estimated the center of the 3D pupil by unprojecting points from the extracted 2D pupil contour under the assumption of refraction occurring at corneal interfaces. As the radius of the unprojected pupil is estimated given measurements from the 2D image pupil, their method is an approximate solution. Hennessy et al. proposed a similar approximation, albeit for a single camera and multiple glints [Hennessey et al. 2006]. Shi and Liu used a pair of calibrated stereo cameras and multiple glints in their gaze-estimation system [Shih and Liu 2004]. However, they assumed the image of the center of the 3D pupil to be the center of the 2D pupil image; but given effects of refraction and perspective projection this is only an approximation. The same holds true for the system proposed in [Guestrin and Eizenman 2006]. The method proposed by Chen et al. was based on the simplifying assumption that the center of the virtual pupil lies on the optical axis, i.e. the line containing both the center of the 3D pupil and the center of curvature of the cornea [Chen et al. 2008]. In summary, while the importance of refraction is well-known, existing methods either disregarded or merely approximated the effects of refraction. Here, we present a novel approach that does not rely on simplifying assumptions with regard to refraction effects implicated by the employed eye model.

2.4 Refraction-induced errors in glint-based remote eye trackers

Villanueva et al. studied the error in angular gaze-estimation accuracy resulting from disregarding/approximating corneal refraction [Villanueva and Cabeza 2008]. Their results showed that disregarding refraction entirely leads to errors as large as several degrees.

Employing approximative approaches for dealing with corneal refraction, such as developed in [Shih and Liu 2004] or [Guestrin and Eizenman 2006], still leads to systematic errors up to almost one degree. Recently, Barsingerhorn et al. presented a simulation study with regard to refraction effects in remote eye trackers using a pair of stereo cameras and two IR light sources [Barsingerhorn et al. 2017]. Their analysis showed that deviations in gaze predictions of up to one degree are to be expected from the standard simplifying assumptions usually made with regard to eye geometry and physiology. In this study, in contrast, we are concerned with gaze estimation in a mobile setting. In particular, we investigate the effects of refraction in a system that does not use glints for determining fundamental model parameters such as eyeball position.

3 METHOD

We propose a novel approach to single camera, glint-free 3D eye model fitting including corneal refraction. After a short summary of the model proposed in [Świrski and Dodgson 2013], we describe the geometric eye model employed in our work, the proposed optimization procedure, and how gaze estimation is performed based on the optimized model.

3.1 The Świrski model

The approach proposed by [Świrski and Dodgson 2013] is based on two assumptions: (i) the apparent pupil contour in a 2D eye image is the perspective projection of a circular 3D pupil which is tangent to an eye sphere of fixed radius, R, (ii) the center of the eye sphere is stationary over time. The first assumption implies that refraction effects are not taken into account. The second assumption is fulfilled e.g. when considering head-mounted eye trackers.

In this approach, changes in gaze direction correspond to movements of the circular pupil with center point P lying on the surface of the eye sphere. At each point in time, the state of the eye model is thus determined by (i) the center of the eye sphere, $E = (s_x, s_y, s_z)$, and a state vector, $v = (\phi, \theta, r)$, comprising the current radius r of the pupil and the spherical coordinates ϕ and θ of the normalized vector $P - E/\|P - E\|_2$, i.e the vector pointing from the eye sphere center to the pupil center. Here, $\|.\|_2$ denotes the ℓ_2 norm. Since P is uniquely determined given v, in the following we will for simplicity refer to v also as the 3D pupil circle.

Given a set of N eye images, recorded over a period of time, pupil contours are extracted from each image by means of an automatic pupil extraction algorithm [Świrski et al. 2012], leading to sets of two-dimensional contour edges $\mathcal{E}_i = \{e_{ij} \text{ with } j = 1, \ldots, M_i\}$, where the subscript ranges over the number of recorded images. The algorithm presented by [Świrski and Dodgson 2013] uses a joint nonlinear optimization of the parameters E and v_i.

In a first step, a non-iterative initialization of all fit parameters is performed. Note, as the problem is only determined up to a global scale due to the properties of perspective projections, the radius of the eye sphere can be fixed at an arbitrary value. For details about the initialization procedure we refer the reader to the original publication [Świrski and Dodgson 2013].

In a second step, a nonlinear optimization is carried out, minimizing the reprojection error of all 3D pupil circles v_i. More specifically,

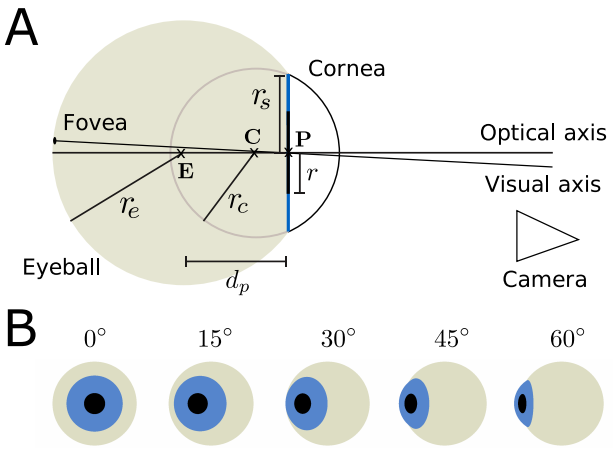

Figure 2: (A) Schematic of the simplified Le Grand eye model. (B) Examples of synthetic eye images used for evaluation. The eye was positioned in front of the camera at a distance of 35mm along the positive z-axis. From left to right, the eye is rotated around the y-axis by an angle as indicated.

each 3D pupil circle v_i, upon perspective projection to the 2D image plane, defines a 2D ellipse, referred to as ℓ_i. The cost function minimized is denoted by Λ and is given by

$$\Lambda(E, v_1, \ldots, v_N) = \sum_i \sum_j d(e_{ij}, \ell_i)^2, \quad (1)$$

where $d(e_{ij}, \ell_i)$ denotes the distance of the pupil contour edge e_{ij} to the ellipse ℓ_i. In order to reduce computation time, [Świrski and Dodgson 2013] chose to employ an approximate signed distance function d, with $|d|$ for ellipses with low eccentricity being close to the usual Euclidean distance of the edge e_{ij} to the ellipse ℓ_i.

3.2 3D eye model

We base our work on the Le Grand eye model [Le Grand 1957] which stipulates the eye to consist of two partial spheres (see Figure 2A): the eyeball and the cornea. The eyeball has its center in $E = (s_x, s_y, s_z)$ and radius of curvature r_e. It is covered with the opaque sclera. The transparent cornea has its center in C and radius of curvature r_c. Note, for simplicity it is assumed that the cornea and the aqueous humor form a homogeneous medium with an effective refractive index n=1.3375 [Guestrin and Eizenman 2006]. The iris and pupil are two concentric circles with normals parallel to $\overline{EC}$ and radius r_s and r, respectively. Their common center lies at a distance $d_p = \sqrt{r_e^2 - r_s^2}$ from E along the direction from E to C. The optical axis is defined as the line through E and P, the visual axis as the line through the fovea and P. The angular offset between the visual and the optical axis can be obtained by a one-point calibration procedure

In the following, we will be concerned with the estimation of the optical axis. Note, all coordinates and measurements are expressed in camera coordinates. In particular, we assume the camera to be placed in the origin of a right-handed coordinate system and to face into the positive z-direction. The state of the model, similar to the one employed by Świrski et al., at any given time is determined by

Table 1: Eye model parameters used in the generation and analysis of synthetic and real eye images.

Parameter name	Symbol	Value	Ref.
Eyeball radius	r_e	12 mm	[Bekerman et al. 2014]
Cornea radius	r_c	7.8 mm	[Guestrin and Eizenman 2006]
Iris radius	r_s	6.0 mm	[Gross 2008]
Pupil radius	r	1-4 mm	[Gross 2008]
Effective corneal refractive index	n	1.3375	[Guestrin and Eizenman 2006]

the eyeball center E and a state vector, $v = (\phi, \theta, r)$, which is given by the spherical coordinates of the normalized vector pointing from E into the direction P, i.e. the optical axis, and the radius of the pupil. We will refer to ϕ and θ as gaze angles. In some cases, we will also refer to the angle between the optical axis and the negative z-axis as gaze angle. In order to assure that $\phi = \theta = 0$ corresponds to a zero gaze angle in the second sense, we adopt the following convention; given the normalized gaze vector $(P-E)/\|P-E\|_2 = (\mathrm{gv}_x, \mathrm{gv}_y, \mathrm{gv}_z)$, we define $(\phi, \theta) = (\arctan(\mathrm{gv}_z/\mathrm{gv}_x) + \pi/2, \arccos(\mathrm{gv}_y) - \pi/2)$.

The parameters used are given in Table 1. We generate synthetic images of the Le Grand eye model at 640x480 pixels resolution by means of a ray-tracing pipeline (see Figure 2B), in which we assume the camera to have a focal length of f=620 pixels.

3.3 Model optimization

Similar to the approach proposed in [Świrski and Dodgson 2013], we define a cost function, denoted by Λ_{ref}, to optimize the parameters E and v_i, based on the sets of edge contours $\mathcal{E}_i$ extracted from a series of eye images. As no closed-form solution exists for calculating the refracted 2D pupil contours, instead of minimizing the reprojection error, we measure model fitness by calculating suitable residuals for unprojections $\Theta(e_{ij})$ of the contour edges e_{ij} in 3D space.

The proposed function Θ is based on an inverse ray-tracing approach. It is designed (i) to account for refraction of rays intersecting the cornea and (ii) to provide viable information about model fitness when rays are not intersecting the cornea. To this end, for each edge e_{ij} we consider the ray ρ_i, emanating from the camera origin and passing through $\hat{e}_{ij}$, where $\hat{e}_{ij}$ is a 3D position of the edge e_{ij} obtained from the projection matrix of the employed camera. We thus define $\rho_i(t) = t \cdot \hat{e}_{ij}$ for $t \in \mathbb{R}$.

We distinguish three cases, depending on the existence of intersections of the ray with the current eye model. We discuss these three cases in turn.

Case I. The first intersection with the eye model along the ray is with the corneal surface. We denote this intersection point by ρ_0. The law of refraction states that the normal $c_\perp$ of the corneal surface at ρ_0, the direction of the incident ray, $\hat{e}_{ij}$, and the direction of the refracted ray, $\hat{e}_{ij}^{\mathrm{ref}}$, are coplanar and that furthermore

$$\hat{e}_{ij}^{\mathrm{ref}} = \frac{n}{n_{\mathrm{air}}}\left(\hat{e}_{ij} - \left(\langle \hat{e}_{ij}, c_\perp \rangle + \sqrt{\frac{n^2}{n_{\mathrm{air}}^2} - 1 + \langle \hat{e}_{ij}, c_\perp \rangle^2}\right) c_\perp\right), \quad (2)$$

where $\langle .,. \rangle$ denotes scalar multiplication and n_{air}=1.0 is the refractive index of air. We define the refracted ray ρ^{ref} as $\rho^{\mathrm{ref}}(t) = \rho_0 + t \cdot \hat{e}_{ij}^{\mathrm{ref}}$ with $t \in \mathbb{R}$ and calculate its intersection with the plane passing through the center of the 3D pupil circle v_i and orthogonal to the optical axis, i.e. the 3D plane that contains the pupil with index i. We denote this intersection by ρ_0^{ref} and define

$$\Theta(e_{ij}) = \rho_0^{\mathrm{ref}}. \quad (3)$$

Case II. The first intersection with the eye model along the ray is with the eyeball. We denote this intersection ρ_0 and define

$$\Theta(e_{ij}) = \rho_0. \quad (4)$$

Case III. There is no intersection with either the eyeball nor the cornea. In this case, we determine $t_0 \in \mathbb{R}$ such that the distance between $\rho(t_0)$ and E is minimal and define $\Theta(e_{ij}) = \rho(t_0)$.

Given the unprojection function Θ, we further define a residual function, κ, as

$$\kappa(e_{ij}) = \begin{cases} |r_i - \|\Theta(e_{ij}) - P_i\|_2|, & \text{if } \Theta(e_{ij}) \text{ is case } \mathbf{I} \\ \|\Theta(e_{ij}) - P_i\|_2, & \text{if } \Theta(e_{ij}) \text{ is case } \mathbf{II} \\ \|\Theta(e_{ij}) - E\|_2 + d_p, & \text{if } \Theta(e_{ij}) \text{ is case } \mathbf{III}. \end{cases} \quad (5)$$

Given our definition of κ, a brief discussion of the three cases defined above is possible. Note that case III encodes a situation, in which the eye model is far away from its correct position. In this case, shifting the eye center E will reduce the residual of an edge e_{ij}, while changes in pupil parameters $v_i = (\phi_i, \theta_i, r_i)$ lead to no change in $\kappa(e_{ij})$. In contrast, case II corresponds to a situation in which the eye model is already close to its optimal position, such that adjustments to E, ϕ_i, and θ_i will reduce the residual $\kappa(e_{ij})$. Finally, case I implies that the pose of the eyeball is such that the ray ρ_i corresponding to e_{ij} intersects the cornea. In this case, we penalize the distance to the actual 3D pupil circle v_i. The residual consequently depends on E and all parameters in v_i.

Note, the addition of d_p in case III assures that $\kappa(e_{ij})$ is always smaller when case II is fulfilled compared to when case III is realized.

We can now define the cost function Λ_{ref} as

$$\Lambda_{\mathrm{ref}}(E, v_1, \ldots, v_N) = \sum_i \frac{1}{\|P_i\|^2} \sum_j \kappa(e_{ij})^2 + \lambda \|\Theta(c_i) - P_i\|^2, \quad (6)$$

where c_i denotes the center point of an ellipse that is fit to the edges $\mathcal{E}_i$ and λ is a parameter for setting the relative strength of the second summand.

Note that the squared residuals belonging to the i-th observation are scaled by the squared inverse distance of P_i to the origin. This scaling is necessary as otherwise a reduction of residuals could trivially be achieved by moving E towards the camera and shrinking all radii r_i to zero.

The second summand enforces that the unprojection $\Theta(c_i)$ of the fitted ellipse center c_i is equal to the pupil center P_i. This constitutes an approximation [Villanueva and Cabeza 2008]. We have observed, however, that in a first step minimizing Λ_{ref} setting λ=400.0, i.e. to a high value, and finalizing the optimization in a second step, after setting λ=0, prevents from undesired edge cases of wrongly fitted pupil contours. Note, the final result is not subject to the above approximative constraint.

When optimizing the parameters of our eye model for a given set of images, for simplicity we employ the same initialization as proposed in [Świrski and Dodgson 2013]. While their initialization

procedure ignores refraction effects, the obtained initialization values still provide a reasonable starting point for the minimization of Λ_{ref}. As in [Świrski and Dodgson 2013], we have implemented our model using the CERES solver [Agarwal et al. 2010], in particular making use of its automatic differentiation capabilities.

3.4 Gaze estimation

In the preceeding subsections, we have shown how the Świrski model and ours can be fit to a set of pupil observations. We now extend both approaches to gaze-estimation systems. Note, Świrski et al. do not discuss how to employ the optimized model for gaze estimation based on additional images. However, the method proposed here is implicitly contained in their initialization procedure.

3.4.1 Gaze-estimation using the Świrski model. Let us assume the parameters E and v_i have been optimized in order to minimize the cost function Λ. To extract a gaze vector from a new observation, first an ellipse ℓ is fit to the contour edges $\mathcal{E}$. Unprojecting this ellipse assuming a full perspective projection by a pinhole camera model and fixing an arbitrary positive radius, results in two 3D circles [Safaee-Rad et al. 1992; Świrski and Dodgson 2013; Wang et al. 2003]. By projecting the normal of each circle to the image plane and checking which of the two resulting vectors points away from the projected eyeball center, one can disambiguate the unprojected circle that is consistent with the current gaze direction [Świrski and Dodgson 2013]. We consider a ray emanating from the origin and passing through the center of the chosen unprojected circle and determine the intersection point with the eye sphere that is closest to the camera. Should no intersection exist, we determine the point on the eye sphere which is closest to the ray. This point will serve as an estimate of the location of the pupil center P and thus determines the gaze direction. By scaling the radius of the unprojected circle appropriately, we also obtain an estimate for the pupil radius [Świrski and Dodgson 2013].

3.4.2 Gaze-estimation using our model. Within the framework of our approach, a gaze estimate is obtained by another minimization. To this end, let us assume that E and v_i have been optimized in order to minimize the cost function Λ_{ref}. In particular, we denote the optimized eyeball center as E_{opt}. Given a new observation with contour $\mathcal{E}$, we define

$$v_{\text{opt}} = \text{argmin}_v \Lambda_{\text{ref}}(E = E_{\text{opt}}, v), \tag{7}$$

i.e. the gaze estimation v_{opt} is obtained by minimizing Λ_{ref} under the constraint that the eyeball center is fixed to the previously established position E_{opt}.

4 EVALUATION

In this section, we present a comparative evaluation of the performance of the Świrski model and the one proposed in this work. To this end, we first use synthetic eye images generated by our ray-tracing pipeline (see Figure 2B for examples). We then also present results obtained on real eye images (see Section 4.5). In order to make the two approaches amenable to a direct quantitative comparison when it comes to estimates of eye sphere/eyeball centre E and pupil radius r, in the Świrski model we chose $R=\sqrt{r_e^2-r_s^2}$. Note that this choice made the eye sphere in the Świrski model tangent to 3D pupil circles in our model, as the same position E was assumed.

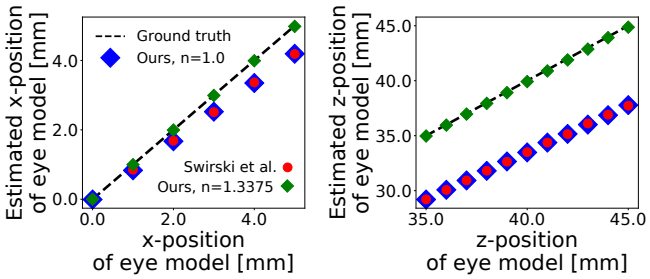

Figure 3: Estimates of eye-model position. (A) Estimated x-position of the respective eye models as a function of the ground truth value of x. (B) The same evaluation as in A, this time shifting the eyeball in z-direction (x=y=0mm).

4.1 Eye model position

We first analyzed whether both approaches can detect the correct position of the eyeball based on a series of images. The accurate determination of the position of the eyeball in 3D is for example of importance for the calibration of optical see-through head-mounted displays [Itoh and Klinker 2014].

We generated a grid of 400 eye images with the eyeball positioned at E=$(x, 0\text{mm}, 40\text{mm})$ with gaze angles ϕ and θ spanning ±40 degrees and varying x between 0mm and 5mm. The pupil radius was fixed at r=2mm. For each x-position, we randomly chose 25 of those images and optimized both the Świrski and our model given the cost functions in Section 3. For our model, we performed the optimization (i) with n=1.3375, i.e. with a refractive cornea, and (ii) with n=1.0, i.e. disregarding refraction effects. The second choice was made to ensure the Świrski model and our approach were equivalent in this case. We repeated this procedure ten times. The results of this analysis are shown in Figure 3.

As can be seen from the figure, the Świrski model and our model with n=1.0 indeed give the same results. However, while for x=0mm all approaches estimate the correct value of x, the error for the Świrski model increases to ≈0.8mm, i.e. ≈15%, when x=5mm (see Figure 3A). Our approach (n=1.3375), in contrast, gives correct estimates for the position x within line width.

We also varied the z-position of the eyeball, while keeping x=y=0mm fixed. While our model with n=1.3375 correctly reproduces the ground-truth values within line width, the Świrski model underestimates the eyeball position by ≈6mm, i.e. ≈17%, over the tested range of z-values (see Figure 3B).

In summary, our data shows that by ignoring refraction the Świrski model, indeed, incurs a systematic error in 3D eyeball position estimates. Also, our optimization approach successfully reproduces the effects introduced by refraction within our synthetic imaging pipeline.

4.2 Angular gaze-estimation accuracy

Next, we evaluated the accuracy of gaze-direction estimates obtained by using the approaches described in Section 3.4. To this end, we assumed the eyeball to be fixed at E=(0mm, 0mm, 35mm) and, as before, generated 400 images on a regular grid, this time spanning ±50 degrees both for ϕ and θ. The pupil radius was again fixed at

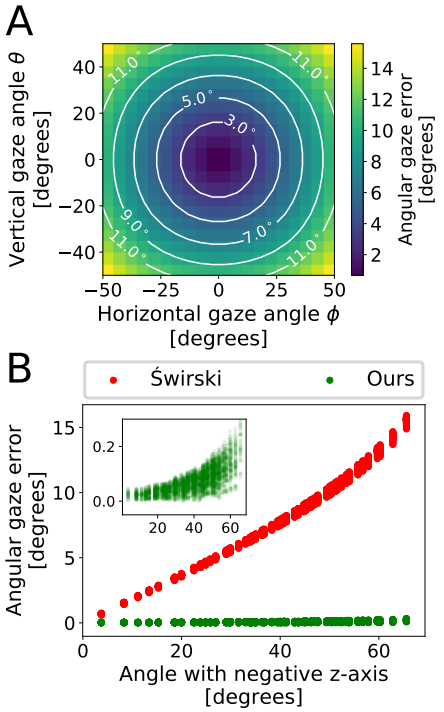

Figure 4: Angular errors in gaze-direction estimates. (A) Average heat map of angular gaze-estimation error in the Świrski model as a function of ϕ and θ. Contour lines indicate levels of equal angular gaze-estimation error. (B) Data shown in A as a function of gaze angle expressed as the angle of the optical axis with the negative z-axis (red dots). We compare to results obtained with the approach proposed in this work (green dots). The inset shows the same data with a rescaled y-axis (same units as in the main panel).

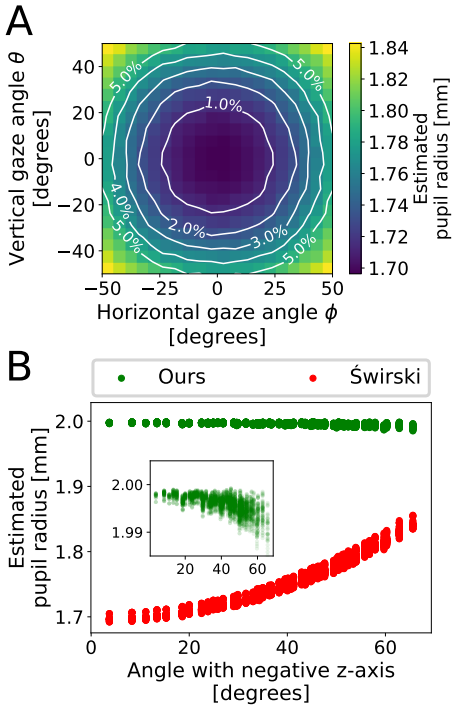

Figure 5: Pupil radius estimates. (A) Average heat meap of pupil-radius estimates obtained with the Świrski model as a function of gaze angles ϕ and θ. The ground-truth value is r=2mm. Contour lines indicate relative changes with respect to the value obtained for ϕ=θ=0. (B) The individual data shown in A as a function of gaze angle expressed as the angle of the optical axis with the negative z-axis (red dots). We compare to results obtained with the approach proposed in this work (green dots). The inset shows the same data with the y-axis rescaled (same units as in the main panel).

r=2mm. We optimized both the Świrski model as well as our model with n=1.3375 based on 25 randomly chosen images. Based on the optimal model parameters, we predicted the gaze direction for all 400 images and calculated the respective angular errors in gaze direction. We repeated this procedure for ten runs and averaged the results. In Figure 4A we show a heat map with overlaid contour lines for the results obtained for the Świrski model. Within numerical accuracy, the gaze prediction has an error of zero degrees at a gaze angle of zero degrees. However, with increasing gaze angle the gaze-estimation error increases up to ≈15 degrees at 50 degree gaze angle. To compare these results with the estimates obtained from our approach, we plot gaze-prediction errors as a function of the angle of the optical axis with the negative z-axis (see Figure 4B). As can be seen from the figure, errors in our approach for all gaze angles are close to zero. There is a systematic shift towards slightly higher errors (≤0.2 degrees) with increasing gaze angle (see inset), which however is most likely due to discretization effects in the image generation pipeline.

This data shows that the Świrski model is affected by a systematic error in predicted gaze angle. Again, our approach successfully predicts the gaze-directions sampled in the synthetic image data.

4.3 Pupil size estimates

We also investigated the accuracy of pupil radius estimates. Pupillometry data, as e.g. the measurement of pupil radii over time, has been shown to contain viable information with regard to e.g. the mental load of a subject [Sirois and Brisson 2014].

Using the gaze estimation results discussed in the last section, in Figure 5A we plot the average pupil radius estimate obtained from the Świrski model as a function of the gaze angles ϕ and θ. We find that while the actual pupil radius used for image generation was r=2mm, the radii estimated by the Świrski model were lower than r≈1.85mm for all gaze angles. However, the data also shows that the pupil-radius estimate is gaze-angle dependent, with radius estimates being smaller for smaller gaze angles. The minimal value found for a gaze direction directly facing the camera is r≈1.70mm. The variation over the whole gaze range thus corresponds to a relative change of ≈5% (see also contours in Figure 5A), i.e. is on the order of response amplitudes in pupil radius measured in experiments in the cognitive sciences [Gagl et al. 2011; Sirois and Brisson 2014]. In Figure 5A we compare the radius estimates obtained with our method to the results generated with the Świrksi model by

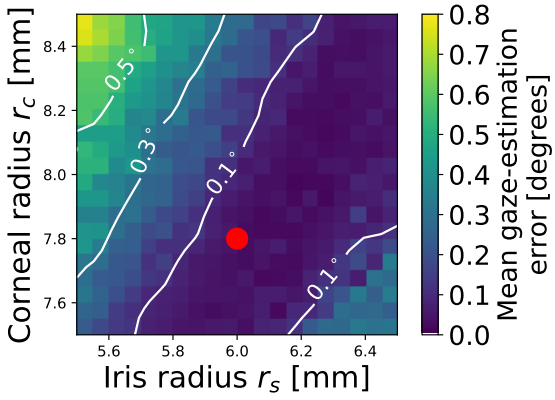

Figure 6: Sensitivity to physiological parameters. Heat map of mean gaze-estimation error as a function of r_c and r_s. Contour lines show levels of the mean gaze-estimation error. The red disk indicates standard parameters (see Table 1).

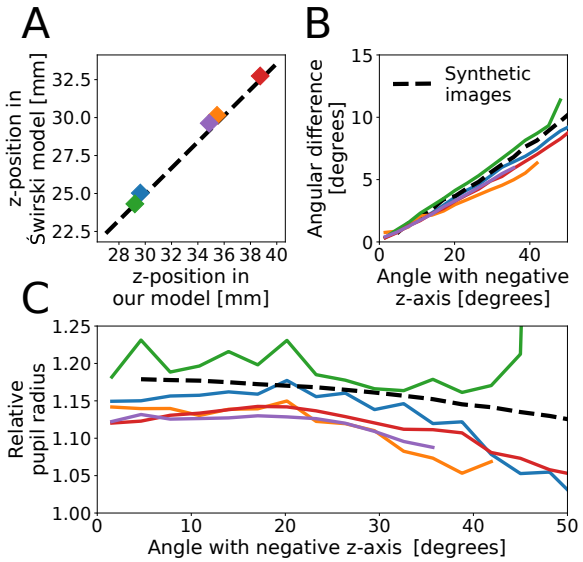

Figure 7: Analysis of real eye images. (A) Average z-position estimated by the Świrski model as a function of the respective value obtained with our approach. (B) Average angular difference between gaze predictions from the Świrski model and our approach as a function of gaze angle with the negative z-axis (see text for details). (C) Average ratio of pupil radii from our approach and the Świrski model. The y-axis is as in B. Colors indicate different subjects. Dashed lines show predictions derived from the results on synthetic images.

plotting each as a function of the angle between the optical axis and the negative z-axis. Our model gives results that are within 0.01mm from the ground-truth value of r=2mm for all gaze angles. A slight systematic increase of errors with increasing gaze angle can be discerned, which again is likely due to discretization effects.

These results show that also with regard to pupil radius estimates, the Świrski model exhibits systematic and gaze-angle dependent errors. Our approach on the other hand provides good estimates for all sample images considered.

4.4 Sensitivity to physiological parameters

So far, we considered physiological parameters within our model to be fixed. Both the corneal radius r_c and the iris radius r_s, however, exhibit variations across the population [Mashige 2013]. We thus furthermore quantified the sensitivity of our method to changes in these parameters. To this end, we generated 400 images on a regular grid spanning ±30 degrees both for ϕ and θ, with eye model parameters given in Table 1. The eyeball was positioned at E=(0mm, 0mm, 35mm) and r=2mm. After choosing values for r_c and r_s, we optimized our model on 25 randomly chosen images. We define the mean gaze-estimation error for a choice of r_c and r_s as the average over all errors on the grid of synthetic images under the assumption that gaze angles, measured as the angle of the optical axis with negative z-axis, are uniformly distributed. We show a heat map of the mean gaze-estimation error as a function of r_c and r_s in Figure 6. These results show that the mean gaze-estimation error stays below 0.3 degrees for most of the parameter choices sampled. Only for small iris radii and large corneal radii, mean gaze-estimation errors above 0.5 degrees are observed. Standard parameters are marked by a red disk and indeed lie within a region exhibiting minimal errors. Also other combinations of r_c and r_s lead to very accurate predictions by our model, with larger iris radii requiring larger corneal radii for best performances.

This data indicates that the gaze-estimation accuracy of our model, while depending on r_c and r_s, is robust with respect to the exact choice of these parameters.

4.5 Real eye images

In order to verify that our approach is also feasible for the analysis of real world data, we recorded eye images of n=5 subjects using the commercially available head-mounted eye tracking headset by Pupil Labs [Kassner et al. 2014]. To mimic our simulation study, for each subject the eye camera was positioned centrally in front of the left eye. Subjects were instructed to gaze freely for about one minute while evenly scanning their natural gaze range. Pupil detection was performed offline using Pupil Capture, the open source software developed by Pupil Labs. For each frame it resulted in a set $\mathcal{E}$ of contour edges and was associated with a confidence value, C, reported by the software. Only frames with a high confidence value were selected ($C \geq 0.98$). For each subject, the Świrksi and our model (with parameters as in Table 1) were optimized independently for 25 times based on 50 randomly selected pupil contours. In Figure 7A the average z-position estimated by the Świrksi model is shown as a function of the respective result obtained with our approach. Note, both the x- and y-position were close to zero in all cases due to the central position of the camera (data not shown). Performing a linear fit to the data contained in Figure 3B, a theoretical prediction for the observed relation was derived (dashed black line). The data shown in Figure 7A is quantitatively consistent with this prediction.

Using the respective eyeball positions reported in Figure 7A, gaze estimates were extracted from all high-confidence pupil observations, both with the Świrksi model and ours. In Figure 7B the

average pairwise angular difference of the two resulting gaze directions is shown as a function of the gaze angle with the negative z-axis as obtained from the gaze direction estimated with our approach. The data was averaged in bins of three degrees and shows that the average angular difference between the two approaches increases from ≈0 degrees at zero degrees gaze angle to ≈10 degrees at ≈50 degrees gaze angle. We find that the observed differences are quantitatively consistent with the results obtained on synthetic images (red dots in Figure 4B are shown here as a dashed line). Small deviations are likely in part due to variations in physiological eye parameters and the different z-positions of the eyeball as compared to the setup analyzed in Figure 4B, where only one eyeball position was considered.

Last, the relation of pupil radii obtained from the Świrski model and our approach was investigated. Note, as pupil radii were not constant over the course of the recording, in Figure 7C relative pupil radii are shown. More specifically, the average ratio of the pupil radius as obtained by our approach and the pupil radius estimated by the Świrski model is shown as a function of the gaze angle with respect to the negative z-axis (definition as in panel A). Data was averaged in bins of three degrees. Over the whole range of gaze angles, the Świrksi model predicts a smaller pupil radius as compared to our approach. However, this effect is attenuated with increasing gaze angle. This behavior is consistent with the results presented in Figure 5B. The ratio of pupil radii as calculated from the data reported in Figure 5B (green vs. red points) is shown as a dashed line. The results obtained from real eye images clearly follow the predicted trend. Deviations, as before, are in part likely due to variations in physiological parameters and different z-positions as compared to the setup discussed in Figure 5B.

Overall, these results show that our approach is suitable also for the analysis of real eye images and that the differences observed between the Świrski model and our approach are consistent with the predictions derived from our simulation study.

5 DISCUSSION

In this work we proposed a novel glint-free approach to 3D eye model fitting solely based on a set of pupil contours. Most importantly, the model also accounts for refraction effects occurring at corneal interfaces. To this end, we proposed a cost function for measuring model fitness. By unprojecting 2D pupil contour edges to suitable points in 3D space, the distance to defining points of the eye model can be evaluated. We presented a detailed comparison of our approach to the Świrski model [Świrski and Dodgson 2013] that does not include refraction modelling. We showed that the Świrski model therefore suffers from systematic errors with regard to eye sphere position, gaze direction, and pupil radius estimates. In contrast, our approach provides accurate estimates for all of the above observables up to the limits in terms of numerical accuracy and discretization effects inherent to our ray-tracing pipeline. By analyzing real eye images, we showed that our approach is also applicable in a real world setting. In particular, the observed differences between the Świrski and our approach were quantitatively consistent with the predictions derived from our simulation study.

One of the key challenges specifically in head-mounted eye tracking are unavoidable movements of the headset relative to the user's head, so-called headset slippage. Headset slippage can occur after only a few minutes of use and has been the subject of recent work [Sugano and Bulling 2015]. Villanueva et al. showed that in a remote glint-based setting hardware calibration can effectively compensate for systematic errors incurred by ignoring refraction [Villanueva and Cabeza 2008]. However, only at the cost of an increased sensitivity to changes of the hardware setup, as e.g. introduced by head movements with respect to the remote camera and light setup. Accounting for refraction effects thus also has potential for making head-mounted eye trackers more robust to headset slippage. We plan to explore this in future work.

Despite its potential for glint-free gaze estimation, particularly in head-mounted settings, our model also has a few limitations. First, in this work we were mainly interested in shedding light on deterministic effects differentiating the Świrski model and our novel approach to contour-based eye tracking. Therefore, most of our evaluations are based on synthetic eye images. While our results show the feasibility of our approach for the analysis of real eye images, the recording of these images was still controlled. It is therefore an important challenge for future work to study and, if required, improve the robustness of the proposed optimization procedure in the presence of real-world noise sources, such as imaging noise, pupil detection errors, and slippage events, possibly occurring on short time scales.

Second, we have opted for a model of ocular optics that constitutes a simplification of the human eye. In a remote glint-based setting, [Barsingerhorn et al. 2017] recently showed that including asphericities and position-dependent radii of curvature into corneal modelling can have repercussions for the qualitative nature and quantitative extent of refraction effects. A more faithful model of refraction effects occurring at corneal interfaces certainly is of interest in the setup considered here. However, by introducing new person specific parameters such approaches also potentially introduce additional error sources. It will be an exciting direction for further research to map out the effects, potential benefits, and drawbacks of more realistic eye models. Note, since our optimization approach is not dependent on an analytic description of the mapping from 3D pupil circles to 2D pupil images, it lends itself to generalizations with respect to eye models utilized in the future.

6 CONCLUSION

We presented evidence that by ignoring refraction effects, the current state-of-the-art approach to contour-based and glint-free gaze estimation suffers severe limitations. Our simulation study has revealed that systematic and significant errors in eyeball position, gaze direction, and pupil radius estimates result from this simplification. Our approach to 3D eye model optimization accounts for refraction at corneal surfaces by means of a novel cost function. It therefore does not suffer from similar shortcomings and promises to provide accurate estimates for all of the above parameters. The framework described here constitutes a promising platform for further investigations aimed at dissecting the importance of more detailed and realistic eye models in glint-free gaze estimation. We therefore strongly believe our work paves the way for further research geared towards pushing the limits of eye tracking in this challenging setting.

REFERENCES

S. Agarwal, K. Mierle, and Others. 2010. Ceres Solver. http://ceres-solver.org. (2010).

A. D. Barsingerhorn, F. N. Boonstra, and H. H. L. M. Goossens. 2017. Optics of the Human Cornea Influence the Accuracy of Stereo Eye-Tracking Methods: a Simulation Study. *Biomedical Optics Express* 8, 2 (2017), 712–725. https://doi.org/10.1364/BOE.8.000712

I. Bekerman, P. Gottlieb, and M. Vaiman. 2014. Variations in Eyeball Diameters of the Healthy Adults. *Journal of Ophthalmology* 2014, 3 (2014). https://doi.org/10.1155/2014/503645

J. Chen and Q. Ji. 2008. 3D Gaze Estimation With a Single Camera Without IR Illumination. In *ICPR 2008. 19th International Conference on Pattern Recognition*. IEEE, 1–4. https://doi.org/10.1109/ICPR.2008.4761343

J. Chen, Y. Tong, W. Gray, and Q. Ji. 2008. A Robust 3D Eye Gaze Tracking System Using Noise Reduction. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications (ETRA '08)*. ACM, New York, NY, USA, 189–196. https://doi.org/10.1145/1344471.1344518

C. Fedtke, F. Mann, and A. Ho. 2010. The Entrance Pupil of the Human Eye: a Three-Dimensional Model as a Function of Viewing Angle. *Optics Express* 18 (2010), 1–13. https://doi.org/10.1364/OE.18.022364

W. Fuhl, T. Santini, G. Kasneci, W. Rosenstiel, and E. Kasneci. 2017. PupilNet v2.0: Convolutional Neural Networks for CPU Based Real Time Robust Pupil Detection. *CoRR* (2017). https://arxiv.org/abs/1711.00112

W. Fuhl, M. Tonsen, A. Bulling, and E. Kasneci. 2016. Pupil detection for head-mounted eye tracking in the wild: An evaluation of the state of the art. *Springer Machine Vision and Applications* 27, 8 (2016), 1275–1288. https://doi.org/10.1007/s00138-016-0776-4

B. Gagl, S. Hawelka, and F. Hutzler. 2011. Systematic Influence of Gaze Position on Pupil Size Measurement: Analysis and Correction. *Behavior Research Methods* 43, 4 (2011), 1171–1181. https://doi.org/10.3758/s13428-011-0109-5

H. Gross (Ed.). 2008. *Handbook of Optical Systems: Vol. 4 Survey of Optical Instruments*. Wiley-VCH Verlag GmbH+Co. KGaA, Weinheim.

E. D. Guestrin and M. Eizenman. 2006. General Theory of Remote Gaze Estimation Using the Pupil Center and Corneal Reflections. *IEEE Transactions on Biomedical Engineering* 53, 6 (2006), 1124–1133. https://doi.org/10.1109/TBME.2005.863952

D. W. Hansen and Q. Ji. 2010. In the Eye of the Beholder: A Survey of Models for Eyes and Gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, 3 (2010), 478–500. https://doi.org/10.1109/TPAMI.2009.30

C. Hennessey, B. Noureddin, and P. Lawrence. 2006. A Single Camera Eye-Gaze Tracking System With Free Head Motion. In *Proceedings of the 2006 Symposium on Eye Tracking Research & Applications (ETRA '06)*. ACM, 87–94. https://doi.org/10.1145/1117309.1117349

Y. Itoh and G. Klinker. 2014. Interaction-Free Calibration for Optical See-Through Head-Mounted Displays based on 3D Eye Localization. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*. 1–8. https://doi.org/10.1109/3DUI.2014.6798846

A.-H. Javadi, Z. Hakimi, M. Barati, V. Walsh, and L. Tcheang. 2015. SET: a Pupil Detection Method Using Sinusoidal Approximation. *Frontiers in Neuroengineering* 8 (2015). https://doi.org/10.3389/fneng.2015.00004

M. Kassner, W. Patera, and A. 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*. ACM, 1151–1160. https://doi.org/10.1145/2638728.2641695

C.-C. Lai, S.-W. Shih, and Y.-P. Hung. 2015. Hybrid Method for 3-D Gaze Tracking Using Glint and Contour Features. *IEEE Transactions on Circuits and Systems for Video Technology* 25, 1 (2015), 24–37. https://doi.org/10.1109/TCSVT.2014.2329362

Y. Le Grand. 1957. *Light, Color and Vision*. Wiley, New York, NY, USA.

M. Mansouryar, J. Steil, Y. Sugano, and A. Bulling. 2016. 3D Gaze Estimation from 2D Pupil Positions on Monocular Head-Mounted Eye Trackers. In *Proc. of the 9th ACM International Symposium on Eye Tracking Research & Applications (ETRA)*. 197–200. https://doi.org/10.1145/2857491.2857530

K. P. Mashige. 2013. A Review of Corneal Diameter, Curvature, and Thickness Values and Influencing Factors. *The South African Optometrist* (2013), 185–194. https://doi.org/10.4102/aveh.v72i4.58

A. Mayberry, P. Hu, B. Marlin, C. Salthouse, and D. Ganesan. 2014. iShadow: Design of a Wearable, Real-time Mobile Gaze Tracker. In *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*. 82–94. https://doi.org/10.1145/2594368.2594388

T. Ohno, N. Mukawa, and A. Yoshikawa. 2002. FreeGaze: A Gaze Tracking System for Everyday Gaze Interaction. *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications (ETRA '02)* (2002), 1–8. https://doi.org/10.1145/507072.507098

R. Safaee-Rad, I. Tchoukanov, K. C. Smith, and B. Banhabib. 1992. Three-Dimensional Location Estimation of Circular Features for Machine Vision. *IEEE Transactions on Robotics and Automation* 8, 5 (1992), 624–640. https://doi.org/10.1109/70.163786

T. Santini, W. Fuhl, and E. Kasneci. 2018. PuRe: Robust Pupil Detection for Real-Time Pervasive Eye Tracking. *Computer Vision and Image Understanding* (2018). https://doi.org/10.1016/j.cviu.2018.02.002

S. W. Shih and J. Liu. 2004. A Novel Approach to 3-D Gaze Tracking Using Stereo Cameras. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)* 34, 1 (2004), 234–245. https://doi.org/10.1109/TSMCB.2003.811128

S. Sirois and J. Brisson. 2014. Pupillometry. *Wiley Interdisciplinary Reviews: Cognitive Science* 5, 6 (2014), 679–692. https://doi.org/10.1002/wcs.1323

Y. Sugano and A. Bulling. 2015. Self-calibrating Head-mounted Eye Trackers Using Egocentric Visual Saliency. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST)*. 363–372. https://doi.org/10.1145/2807442.2807445

L. Świrski, A. Bulling, and N. Dodgson. 2012. Robust Real-Time Pupil Tracking in Highly Off-Axis Images. In *Proceedings of the Symposium on Eye Tracking Research & Applications (ETRA '12)*. 173–176. https://doi.org/10.1145/2168556.2168585

L. Świrski and N. A. Dodgson. 2013. A Fully-Automatic, Temporal Approach to Single Camera, Glint-Free 3D Eye Model Fitting. In *Proceedings of ECEM 2013*. http://www.cl.cam.ac.uk/research/rainbow/projects/eyemodelfit/

M. Tonsen, J. Steil, Y. Sugano, and A. Bulling. 2017. InvisibleEye: Mobile Eye Tracking Using Multiple Low-Resolution Cameras and Learning-Based Gaze Estimation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* (2017). https://doi.org/10.1145/3130971

M. Tonsen, X. Zhang, Y. Sugano, and A. Bulling. 2016. Labelled pupils in the wild: A dataset for studying pupil detection in unconstrained environments. In *Proc. of the 9th ACM International Symposium on Eye Tracking Research & Applications (ETRA)*. 139–142. https://doi.org/10.1145/2857491.2857520

A. Tsukada and T. Kanade. 2012. Automatic Acquisition of a 3D Eye Model for a Wearable First-Person Vision Device. In *Proceedings of the Symposium on Eye Tracking Research & Applications (ETRA '12)*. https://doi.org/10.1145/2168556.2168597

A. Tsukada, M. Shino, M. Devyver, and T. Kanade. 2011. Illumination-Free Gaze Estimation Method for First-Person Vision Wearable Device. *IEEE International Conference on Computer Vision Workshops* (Nov. 2011), 2085–2091. https://doi.org/10.1109/ICCVW.2011.6130505

A. Villanueva and R. Cabeza. 2008. Evaluation of Corneal Refraction in a Model of a Gaze Tracking System. *IEEE Transactions on Biomedical Engineering* 55, 12 (2008), 2812–2822. https://doi.org/10.1109/TBME.2008.2002152

J.-G. Wang, E. Sung, and R. Venkateswarlu. 2003. Eye Gaze Estimation from a Single Image of One Eye. *Proceedings of the Ninth IEEE Internatinal Conference on Computer Vision (ICCV)* (2003), 1–8. https://doi.org/10.1109/ICCV.2003.1238328