Fabian Leven* and Michael Heizmann

Influence of undetected corneal glints on gaze estimation during manual work

Einfluss undetektierter Cornea-Reflexionen auf die Schätzung der Blickrichtung im Kontext von Handarbeit

DOI 10.1515/teme-2022-0055

Abstract: This work is concerned with model-based remote gaze estimation through monocular video oculography using the pupil center and corneal reflections in the context of manual work. Based on simulations, the influence of undetected corneal glints on the estimated direction of gaze is quantified and discussed.

Keywords: gaze estimation, manual work, corneal reflections

Zusammenfassung: Diese Arbeit befasst sich mit der stationären, monokularen, auf Video-Okulographie basierenden Blickregistrierung zur Schätzung der Blickrichtung im Kontext von Handarbeit. Die Schätzung basiert auf einem geometrischen Modell in Bezug auf das Pupillenzentrum und Cornea-Reflexionen. Anhand von Simulationen werden die Auswirkungen von undetektierten Cornea-Reflexionen auf die geschätzte Blickrichtung quantifiziert und diskutiert.

Schlüsselwörter: Blickrichtungsschätzung, Handarbeit, Cornea-Reflexionen

1 Introduction and related work

This work is in the context of remote gaze estimation through monocular video oculography without any worn sensors (in the following only termed *gaze estimation*). There exists a broad variety of methods for gaze estimation [3]. A class of methods relies on *corneal glints* that result from near-infrared LEDs (*NIR-LEDs*). Several works [2, 4, 6] show, that it is possible to determine the line of sight based on a model of the gaze estimation system (in the following just termed *system*) as well as the human eye in combination with calibration procedures. Using two NIR-LEDs and the two corresponding corneal glints is sufficient to estimate the line of sight. However, deploying more NIR-LEDs and using more corneal glints in the estimation process can decrease the angular error of the estimated line of sight. This has been experimentally shown in [6] where, on average, the angular error has improved by 0.33° when using 4 instead of 2 glints. However, the used experimental setup is not comparable to the particular hardware setup for gaze estimation during manual work that is investigated in this work.

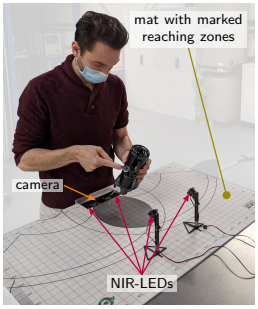
***Corresponding author: Fabian Leven,** Institute of Industrial Information Technology, Karlsruhe Institute of Technology, Karlsruhe, Germany, e-mail: fabian.leven@kit.edu
Michael Heizmann, Institute of Industrial Information Technology, Karlsruhe Institute of Technology, Karlsruhe, Germany

2 Motivation and contribution

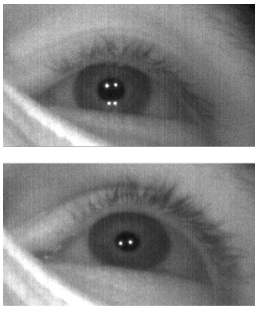
Typical scenarios for systems that rely on corneal glints from NIR-LEDs are in front of a computer screen or in the cockpit of a vehicle. However, in this work, investigations are presented on whether such an approach might be applied to gaze estimation for manual work (such as a disassembly task on a workbench). This work considers one particular, prototypical system with 4 NIR-LEDs and one monochromatic camera mounted on a workbench as shown in Figure 1(a). It is challenging to find a setup geometry that minimizes occlusions by hands, eyelids, and objects and at the same time enables the reflection of light rays originating from the NIR-LEDs at the cornea into the camera. Figure 1(b) shows an image where all 4 corneal glints are visible and an image, where two glints are occluded by the lower eyelid. Based on simulations, it is investigated how the angular error of the estimated line of sight increases due to deviations in the 2D localization of corneal glints if one or more of the corresponding corneal glints are not detected or occluded, respectively.

3 Structure of this article

The analysis in this work is done on simulations. In section 4, the corresponding model is introduced. In section 5,



(a) Hardware setup (adapted from our previous work [1]).



(b) Two image sections that show the left eye.

Fig. 1: The setup that is considered in this work.

the model is fitted to real data to find parameter values that closely resemble the real world gaze estimation system in the context of manual work that is shown in Figure 1(a). In section 6, the found parameter values serve as a basis for a Monte Carlo simulation. In section 7, the line of sight is then estimated from simulated data and compared to the simulation truth to assess the influence of undetected corneal glints on its angular error. In section 8, the results are discussed and an outlook is given.

4 Model for the inference of gaze

The model used in this work is based on the works [2, 4, 6] and adapted to the scene shown in Figure 1(a). In these referenced works, the line of sight was inferred from real data, and e. g., in [2], the angular error was below 1°. Although some deficiencies of the model are known, this confirms that it describes reality reasonably. The model is not explained in detail here. The interested reader is referred to the referenced works [2, 4, 6]. But in short, the model consists of an imaging model, that models the projection of a 3D point in the camera (or world) coordinate system onto a position in a 2D image and an eye model with a spherical cornea to model the reflection of light rays from the NIR-LEDs into the camera. In the subsections 4.1 and 4.2, the parameters of the model are described and in subsection 4.3, the inference of the line of sight from real or simulated data is explained.

4.1 Model parameters

In the following, the relevant input and output parameters of the model are introduced and organized in 4 categories/sets. All 3D coordinates are given in the camera coordinate system.

Category A: Hardware

The parameters of the imaging model describe how a 3D point $[x, y, z]^\top$, $z \neq 0$ in the camera coordinate system is projected onto a 2D position $[u, v]^\top$ in an image coordinate system. The pinhole camera model is used:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} f_x \cdot x \,/\, z + c_x \\ f_y \cdot y \,/\, z + c_y \end{bmatrix} . \tag{1}$$

It has two parameters for the focal length $f_x, f_y > 0$ and two parameters for the optical center of the camera $c_x, c_y > 0$. (As an aside, for practical purposes, also the transformation from the world to the camera coordinate system should be considered, but this is irrelevant for the analysis in this work.) One has to account for non-linear image distortions within a lens model. However, this work is concerned with point-like image features which are assumed to be detected by image-processing algorithms after the image has already been un-distorted.

Further parameters $\mathbf{l}_1, \ldots, \mathbf{l}_4 \in \mathbb{R}^3$ denote the point-like-modeled positions of the 4 NIR-LEDs in the camera coordinate system.

Category B: Person-specific parameters

The human cornea is modeled as a section of the surface of a sphere. In the following, the center of this sphere is termed *corneal center*. The radius $a > 0$ of the said sphere is a person-specific parameter. The optical axis of the eye is understood as the straight line through the corneal center and the center of the pupil. The distance $d > 0$ of the pupil from the corneal center, likewise, is a person-specific parameter. The *visual axis* points along the line of sight. It is modeled as passing through the corneal center. The angular relationship between the optical and the visual axis of the eye can be modeled using two person-specific angles $\alpha, \beta \in [0, 2\pi)$. (The referenced works mention one more person-specific parameter, namely an effective refractive index of the cornea and the aqueous humor. The refraction of light when it passes through the layers of the human eye slightly affects the image of the pupil, but is neglected in this work.)

Category C: Line of sight

The line of sight is modeled as a half-line, that originates at the corneal center $\mathbf{c}$. Furthermore, the visual axis of the eye $\mathbf{v} \in \mathbb{R}^3$ with $\|\mathbf{v}\| = 1$ points along the line of sight. Estimating $\mathbf{c}$ and $\mathbf{v}$ from data may be interpreted as *gaze estimation*.

Category D: 2D positions of image features

For brevity, in the following, the 2D positions of point-like-modeled image features in the coordinate system of an un-distorted image are only termed *image features*. The following image features are considered in the model: The

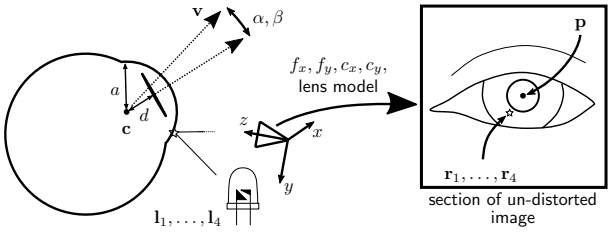


Fig. 2: The model parameters (based on [6]).

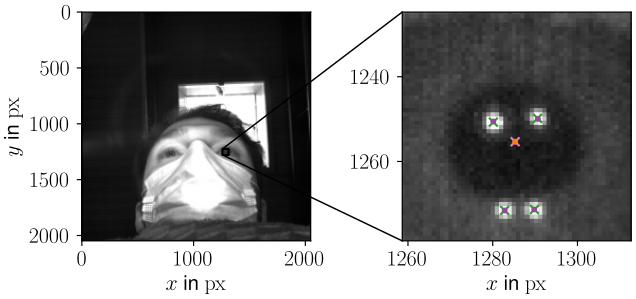


Fig. 3: Result of fitting the model to the measured image features. The measured and modeled pupil center (×, •) as well as the corneal glints (×, •) are shown. (It should be noted that the image has been un-distorted using a calibrated lens model.)

pupil center and the 4 corneal glints $\mathbf{p}, \mathbf{r}_1, \ldots, \mathbf{r}_4 \in \mathbb{R}^2$. In a gaze estimation system, these features may be estimated by image-processing algorithms. In the context of simulations, these features can be calculated from the parameter values of the other three categories (A, B and C).

Figure 2 summarizes all the parameters of the 4 categories:

$$\mathrm{A} = \{f_x, f_y, c_x, c_y, \mathbf{l}_1, \ldots, \mathbf{l}_4\} \,, \qquad \mathrm{B} = \{a, d, \alpha, \beta\} \,,$$
$$\mathrm{C} = \{\mathbf{c}, \mathbf{v}\} \,, \qquad \mathrm{D} = \{\mathbf{p}, \mathbf{r}_1, \ldots, \mathbf{r}_4\} \,.$$

4.2 Parameter notation

It makes sense to distinguish types of parameter values in the remainder of this work, and, therefore, the following symbol decorations are introduced:

The subscript $_\mathrm{t}$ labels the true value of a parameter. Parameter inference from real measurement data aims at determining this value as good as possible, but it remains unknown. However, in the context of parameter inference from simulated data, the true value of a parameter is known.

The subscript $_\mathrm{d}$ (for data) labels either the measured value of a parameter that was determined in some external process, that is not discussed in detail in this work, or the value of a parameter from a simulated measurement.

A hat $\hat{}$ labels the estimated value of a parameter where the estimation process is part of this work, which only applies to the estimated corneal center $\hat{\mathbf{c}}$ and the visual axis $\hat{\mathbf{v}}$.

The subscript $_\mathrm{m}$ labels a function that calculates the value of a parameter in dependence of other parameters according to the introduced model. E. g., if the values of the categories A, B and C are given, the introduced model can be used to calculate the 2D position of the ith corneal glint:

$$\mathbf{r}_{i,\mathrm{m}} \equiv \mathbf{r}_{i,\mathrm{m}}(\mathrm{A}, \mathrm{B}, \mathrm{C}) \,.$$

4.3 Inference of the line of sight

If the values of the categories A, B and D are given, the values of category C (line of sight) can be estimated. This is done in two steps. First, the corneal center is estimated. In this work, the estimated corneal center $\hat{\mathbf{c}}$ is the solution of the following optimization problem, where the euclidean distance of the measured and modeled 2D positions of the corneal glints is minimized:

$$\hat{\mathbf{c}} = \underset{\mathbf{c}\in\mathbb{R}^3}{\arg\min} \sum_{i\in\Omega} \left\| \mathbf{r}_{i,\mathrm{d}} - \mathbf{r}_{i,\mathrm{m}}(\mathrm{A}, \mathrm{B}, \mathbf{c}) \right\| \,, \tag{2}$$

where $\Omega \in \mathcal{P}[\{1,2,3,4\}]$ denotes the set of indexes of the successfully localized corneal glints and $\mathcal{P}[\cdot]$ denotes the power set. For the problem to be well-defined, it must hold $\|\Omega\| \geq 2$ (see [2, 6]). In other words, at least two corneal glints have to be localized successfully. In a second step, the visual axis is estimated as the solution of the following optimization problem, where the euclidean distance of the measured and modeled 2D position of the pupil center is minimized:

$$\hat{\mathbf{v}} = \underset{\mathbf{v}\in\mathbb{R}^3,\, \|\mathbf{v}\|=1}{\arg\min} \left\| \mathbf{p}_\mathrm{d} - \mathbf{p}_\mathrm{m}(\mathrm{A}, \mathrm{B}, \hat{\mathbf{c}}, \mathbf{v}) \right\| \,. \tag{3}$$

5 Finding a simulation truth for gaze estimation during manual work

In the following, the parameter values that closely resemble the scene and setup in the context of manual work as depicted in Figure 1(a) will be determined.

Category A: Hardware

The values of category A are obtained through a cali-

bration procedure of the camera and lens[1] based on a checkerboard pattern. The positions of the 4 NIR-LEDs are obtained using a mirror and an extension of the deflectometric methods presented in [5].

Category B: Person-specific parameters

The values of category B can be determined for an individual person through a calibration procedure. However, in this work, values from literature are used (see [2, 6] and references therein).

Category C: Line of sight

The values of category C change with every camera frame during video oculography. However, in this work, one particular line of sight is considered, namely the one, that corresponds to the scene in Figure 1(a) where the point of regard lies on the screw nut that is loosened. First, the 2D positions of image features of category D were measured, that means extracted from an image, that was taken at the depicted moment. Figure 3 shows this image with an overlay of the features. Now, that the measured values of the categories A, B and D are known, the introduced model can be fitted to these data (according to subsection 4.3) in order to infer the values of category C.

Category D: 2D positions of image features

By fitting the model to the measured values of category D according to the previous sentence, one also obtains modeled values of category D (both, the measured and modeled values, are shown in Figure 3). The latter are used as simulation truth in the remainder of this work. E. g., the simulation truth of the ith corneal glint is given by

$$\mathbf{r}_{i,\mathrm{t}} = \mathbf{r}_{i,\mathrm{m}}(\mathrm{A_t}, \mathrm{B_t}, \mathrm{C_t})\,.$$

Table 1 summarizes all the values of the 4 categories. Furthermore, Figure 4 shows 3D aspects of the model for this choice of parameter values, which may be compared to Figure 1. (Note that only the left eye is considered in this work.)

Table 1: Parameter values that serve as simulation truth. All 3D coordinates are given in the camera coordinate system.

Description	Symbol	Simulation truth
Category A: Hardware (obtained through a setup calibration)		
camera matrix	$f_{x,\mathrm{t}}$	25.27 mm
	$f_{y,\mathrm{t}}$	25.22 mm
	$c_{x,\mathrm{t}}$	5.82 mm
	$c_{y,\mathrm{t}}$	6.14 mm
positions of the NIR-LEDs	$\mathbf{l}_{1,\mathrm{t}}$	$(-0.076\,\mathrm{m}, -0.004\,\mathrm{m}, 0.019\,\mathrm{m})^\top$
	$\mathbf{l}_{2,\mathrm{t}}$	$(\ 0.066\,\mathrm{m}, -0.009\,\mathrm{m}, 0.021\,\mathrm{m})^\top$
	$\mathbf{l}_{3,\mathrm{t}}$	$(-0.106\,\mathrm{m}, -0.379\,\mathrm{m}, 0.217\,\mathrm{m})^\top$
	$\mathbf{l}_{4,\mathrm{t}}$	$(\ 0.083\,\mathrm{m}, -0.375\,\mathrm{m}, 0.226\,\mathrm{m})^\top$
Category B: Person-specific parameters (taken from literature)		
corneal radius	a_t	9.3 mm
distance of pupil from corneal center	d_t	5.1 mm
angular relation of visual and optical axis	α_t	1.5°
	β_t	5.0°
Category C: Line of sight (obtained by fitting the model to data)		
corneal center	$\mathbf{c}_\mathrm{t}$	$(3.35\,\mathrm{cm}, 2.27\,\mathrm{cm}, 67.22\,\mathrm{cm})^\top$
visual axis	$\mathbf{v}_\mathrm{t}$	$(-0.06, -0.50, -0.86)^\top$
Category D: 2D positions of image features (modeled)		
pupil center	$\mathbf{p}_\mathrm{t}$	$(1285.29\,\mathrm{px}, 1255.34\,\mathrm{px})^\top$
corneal glints	$\mathbf{r}_{1,\mathrm{t}}$	$(1282.82\,\mathrm{px}, 1271.57\,\mathrm{px})^\top$
	$\mathbf{r}_{2,\mathrm{t}}$	$(1289.79\,\mathrm{px}, 1271.32\,\mathrm{px})^\top$
	$\mathbf{r}_{3,\mathrm{t}}$	$(1280.13\,\mathrm{px}, 1250.56\,\mathrm{px})^\top$
	$\mathbf{r}_{4,\mathrm{t}}$	$(1290.66\,\mathrm{px}, 1249.95\,\mathrm{px})^\top$

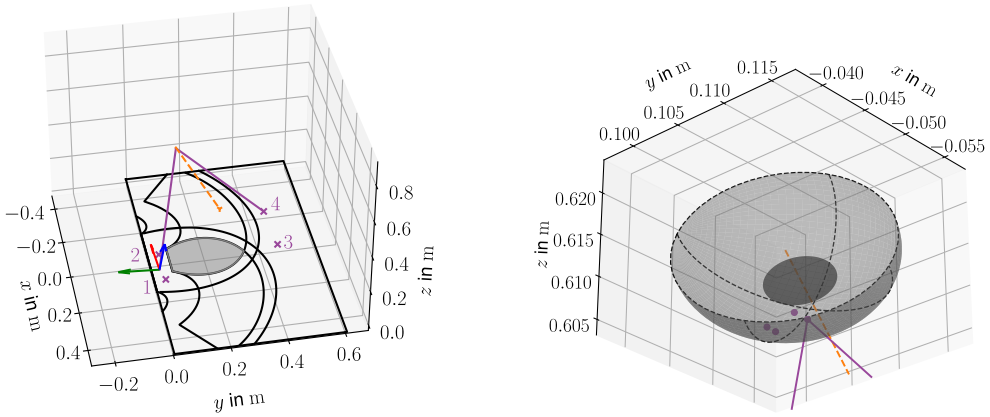
6 Monte Carlo simulation of image features

In this work, the interest lies on the 2D positions of the corneal glints (and the 2D position of the pupil center is not treated in further detail). In a real gaze estimation system, the 2D position of a corneal glint may be determined with a slight random deviation from the true position by an image-processing algorithm. (Systematic deviations are not considered in this work.) The deviation may e. g., stem from image noise or deficiencies of the algorithm. Therefore, the exact nature of the deviation depend´s on the algorithm which means, some assumption has to be made. In this work, the detected 2D position of the ith corneal glint $\mathbf{r}_{i,\mathrm{d}} \in \mathbb{R}^2$ is modeled as random variable, that is drawn from a 2D normal distribution that is centered around the true location $\mathbf{r}_{i,\mathrm{t}}$ and has a diagonal covariance matrix

$$\Sigma = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \tag{4}$$

with $\sigma \geq 0\,\mathrm{px}$. Here, σ may be interpreted as an uncertainty of the localization of corneal glints in an image. Figure 5 shows an overlay of 10 000 simulations for $\sigma = 1\,\mathrm{px}$. Estimating the visual axis from the slightly displaced image features results in an angular error. The estimation of

1 camera product name: `Basler acA2040-90um`, lens product name: `KOWA LM25FC24M`



(a) Overview of the work space.

(b) Close-up of the corneal model.

Fig. 4: Simulation truth on the basis of the scene in Figure 1(a). Subfigure (a) shows an overview of the work space with the point-like-modeled NIR-LEDs (×, 1...4), the camera coordinate system with x-, y- and z-direction (→, →, →) and a light ray (—) that originates from a NIR-LED and is reflected on the corneal surface into the camera. (For better readability, only one ray instead of 4 per eye is shown). Furthermore, the line of sight (– –) is shown. Subfigure (b) shows a close-up of the corneal model with the pupil (slightly transparent black disc) and the corneal surface (slightly transparent gray surface) and the 4 points of reflection (corneal glints) of the light rays from the NIR-LEDs (●). (The coordinate system is the same in the Subfigures (a) and (b) and does not coincide with the camera coordinate system.)

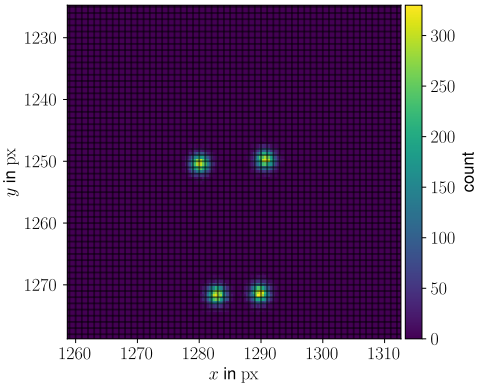


Fig. 5: Histogram of 10 000 simulations (corresponds to 40 000 corneal glints) of simulated 2D positions of the corneal glints $\mathbf{r}_{1,\mathrm{d}}, \dots \mathbf{r}_{4,\mathrm{d}}$ for $\sigma = 1\,\mathrm{px}$. The grid lines are the pixel boundaries (not to be confused with the bin edges of the histogram which are more fine grained). The image section corresponds to the one in Figure 3 on the right.

the visual axis is more robust against such deviations if more corneal glints are available.

7 Assessment of the angular error of the visual axis

According to the previous section 6 one obtains 10 000 simulated measurements of corneal glints. The line of sight (values of category C) is estimated from these simulated corneal glints according to subsection 4.3. In the inference process, the simulation truth is used for all model parameters except for the corneal glints. One may repeat this process for multiple values of σ, see (4). And, in order to asses the influence of undetected corneal glints, one may also repeat this process for different index sets of detected corneal glints Ω, see (2).

The estimated visual axis $\hat{\mathbf{v}}$ can be compared to the true value of the visual axis $\mathbf{v}_\mathrm{t}$ in order to obtain the angular error

$$\phi = \arccos(\hat{\mathbf{v}}^\top \mathbf{v}_\mathrm{t})\,. \tag{5}$$

As the estimation is repeated 10 000 times (for fixed σ and Ω), one obtains a distribution for ϕ. This distribution of the angular error ϕ was obtained for all combinations of $\sigma \in \{0.00\,\mathrm{px}, 0.25\,\mathrm{px}, 0.50\,\mathrm{px}, 1.00\,\mathrm{px}, 2.00\,\mathrm{px}\}$ and $\Omega \in \mathcal{P}[\Omega]$ with $\|\Omega\| \geq 2$. Table 2 shows the median $\mathrm{M}[\phi]$ of the distributions for $\sigma = 0.50\,\mathrm{px}$ and all 11 different choices of Ω. Figure 6 shows selected distributions for 3 different amounts of corneal glints for $\sigma = 0.50\,\mathrm{px}$. Figure 7 shows the medians of the same selection over σ.

8 Discussion and outlook

The following observations can be made:

Table 2: Median of the angular error $\mathrm{M}[\phi]$ in dependence on the index set of detected corneal glints Ω for $\sigma = 0.50\,\mathrm{px}$ and sorted by $\mathrm{M}[\phi]$.

Ω	$\{1,2,3,4\}$	$\{1,2,4\}$	$\{1,2,3\}$	$\{1,3,4\}$	$\{2,3,4\}$	$\{1,4\}$	$\{2,3\}$	$\{1,3\}$	$\{2,4\}$	$\{1,2\}$	$\{3,4\}$
$\mathrm{M}[\phi]$ in °	0.53	0.60	0.60	0.64	0.64	0.73	0.75	0.77	0.77	1.26	1.36

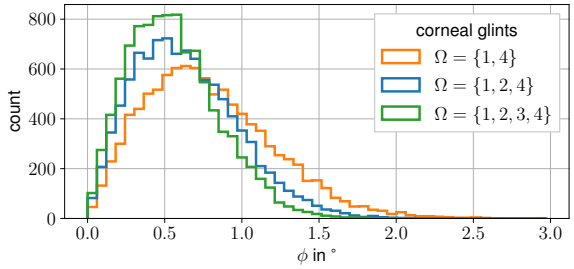


Fig. 6: Distribution of the angular error ϕ for 3 selected index sets of detected corneal glints Ω for $\sigma = 0.50\,\mathrm{px}$. The histograms have 10 000 entries each.

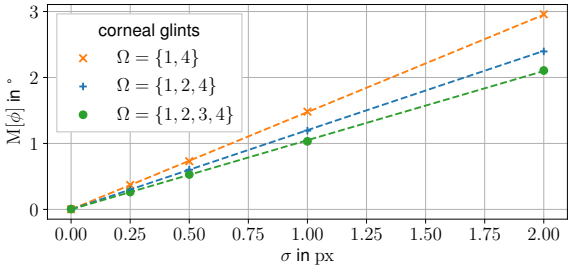


Fig. 7: Median of the angular error of the estimated visual axis $\mathrm{M}[\phi]$ in dependence on σ and Ω.

When only 2 corneal glints are detected, the glints 1 and 2 (NIR-LEDs on the near side of the work space close to the hip) yield better results than glints 3 and 4 (NIR-LEDs on the far side of the work space) while a combination of glints from the near and far side yields the best results (see Table 2). To the best of our knowledge, the considered setup may be regarded as unique concerning the context of manual work and the placement of the two NIR-LEDs near the human hip. The presented results are in favor of this choice of placement.

The distribution of the angular error for fixed σ can span several degrees (see Figure 6), but it is condensed if more corneal glints are detected.

The median angular error depends approximately linearly on σ (see Figure 7). The slope of the linear dependence increases from $\sim 1.1\,°\,\mathrm{px}^{-1}$ if all 4 corneal glints are detected to $\sim 1.5\,°\,\mathrm{px}^{-1}$ if only the glints 1 and 4 are detected.

Concerning the scope of the results, the following has to be considered: The angular error of the estimated line of sight was determined for the hypothetical scenario that the used model only deviates negligibly from reality and all parameters except for the corneal glints can be measured with a negligible uncertainty. Hence, the results may be interpreted as an indicator to what extent the detection and localization of corneal glints contribute to an overall uncertainty of the estimated line of sight in the case of negligible correlations to uncertainties of other parameters. Furthermore, the presented analysis considers one particular scene and may be regarded as a first step in the model-based analysis of uncertainties in gaze estimation in the context of manual work. In the future, it may make sense to examine an extended range of configurations, such as different head or eye poses, and also to include further parameters along with the corneal glints.

Acknowledgment: This work was funded by the Carl Zeiss Foundation within the project AgiProbot.

References

[1] C. R. Dreher, M. Zaremski, F. Leven, D. Schneider, A. Roitberg, R. Stiefelhagen, M. Heizmann, B. Deml and T. Asfour. Erfassung und Interpretation menschlicher Handlungen für die Programmierung von Robotern in der Produktion. *at - Automatisierungstechnik*, 2022. Accepted for publication.

[2] E. D. Guestrin and M. Eizenman. General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Transactions on Biomedical Engineering*, 53(6), 2006.

[3] D. W. Hansen and Q. Ji. In the Eye of the Beholder: A Survey of Models for Eyes and Gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3):478–500, 03 2010. ISSN 1939-3539. 10.1109/TPAMI.2009.30.

[4] S.-W. Shih and J. Liu. A Novel Approach to 3-D Gaze Tracking Using Stereo Cameras. *IEEE transactions on systems, man, and cybernetics.*, 34, 2004. 10.1109/TSMCB.2003.811128.

[5] K. Takahashi, S. Nobuhara and T. Matsuyama. A new mirror-based extrinsic camera calibration using an orthogonality constraint. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, p. 1051–1058, 2012. 10.1109/CVPR.2012.6247783.

[6] A. Villanueva and R. Cabeza. Models for Gaze Tracking Systems. *EURASIP Journal on Image and Video Processing*, 2007(1), 2007. ISSN 1687-5281. 10.1155/2007/23570.