Point of Gaze Estimation through Corneal Surface Reflection in an Active Illumination Environment

Atsushi Nakazawa[1,2,⋆] and Christian Nitschke[1,⋆]

[1] Cybermedia Center, Osaka University
1-32 Machikaneyama, Toyonaka, Osaka 560-0032, Japan
{nakazawa,christian.nitschke}@cmc.osaka-u.ac.jp
[2] PRESTO, Japan Science and Technology Agency (JST)
4-1-8 Honmachi, Kawaguchi, Saitama 332-0012, Japan

Abstract. Eye gaze tracking (EGT) is a common problem with many applications in various fields. While recent methods have achieved improvements in accuracy and usability, current techniques still share several limitations. A major issue is the need for external calibration between the gaze camera system and the scene, which commonly restricts to static planar surfaces and leads to parallax errors. To overcome these issues, the paper proposes a novel scheme that uses the corneal imaging technique to directly analyze reflections from a scene illuminated with structured light. This comprises two major contributions: First, an analytic solution is developed for the forward projection problem to obtain the gaze reflection point (GRP), where light from the point of gaze (PoG) in the scene reflects at the corneal surface into an eye image. We also develop a method to compensate for the individual offset between the optical axis and true visual axis. Second, introducing active coded illumination enables robust and accurate matching at the GRP to obtain the PoG in a scene image, which is the first use of this technique in EGT and corneal reflection analysis. For this purpose, we designed a special high-power IR LED-array projector. Experimental evaluation with a prototype system shows that the proposed scheme achieves considerable accuracy and successfully supports depth-varying environments.

1 Introduction

Eye movements contribute a key part to the interpreting and understanding of a person's wishes, needs, tasks, cognitive processes, affective states, and interpersonal relations. As this information is relevant to a large number of applications in a variety of fields, eye gaze tracking (EGT) is a common problem with a long history [1]. These days, most popular and usable approaches are vision-based techniques [2]. This sort of systems employ an eye camera to capture a

⋆ Both authors contributed equally to this work.

A. Fitzgibbon et al. (Eds.): ECCV 2012, Part II, LNCS 7573, pp. 159–172, 2012.

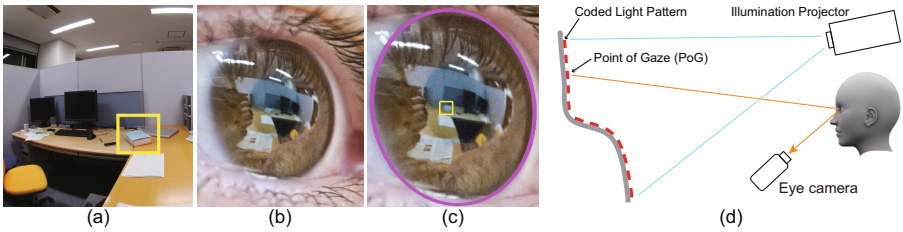


Fig. 1. Corneal reflection and point of gaze (PoG). (a) An environment image showing a book a person is looking at (yellow square). (b) A corresponding corneal reflection image. Note, that this picture is taken in a finely controlled setup, where the camera is placed relatively near to the eye. (c) Estimated iris contour (magenta ellipse) and gaze reflection point (GRP), where light from the PoG is reflected (yellow square). Here, we can see that the GRP coincides with the book. The proposed method analytically obtains the GRP based on a geometric eye model and a reflection model of the corneal surface. (d) A schematic illustration of the PoG estimation approach in an active illumination environment, where the scene is illuminated using coded structured light projection. This allows robust and accurate matching of the GRP in a corneal image to locate the corresponding PoG in a scene image, even from low quality eye images.

close-up view of a person's eye and estimate its pose w.r.t. the camera, then output the point of gaze (PoG) as a location in a planar scene or an image of the environment. For example, the pupil-center/corneal-reflection (PCCR) technique [3] obtains the pose of the eye using the center or contour of the pupil and corneal surface reflections (glints) from point light sources (usually IR LEDs). This technique is the foundation for many current EGT systems; however, it has two major drawbacks: First, mapping the pose of the eye to the PoG requires *system calibration* to obtain the geometric relationship between eye camera and scene. In remote systems, where the relationship is static, calibration determines the pose of the eye camera w.r.t. the scene. In mobile head-mounted systems, where the relationship is dynamic, calibration determines the pose of the eye camera w.r.t. an environment camera that points in the direction of the person's face. In either case, calibrations require a lot of setup effort. Second, due to the fundamental concepts in calculating the PoG, systems potentially suffer from a *parallax issue*. In remote systems the PoG is calculated as the intersection of the gaze ray and a scene model. Since such a model is usually difficult to obtain, systems restrict to simple target scenes like computer screens and walls, or provide functionalities of parallax compensation by manual specification of the scene depth. Moreover, this requires re-calibration when the configuration is changed. In mobile head-mounted systems the PoG is calculated by mapping the gaze ray to an image of the scene assuming a virtual 'calibration plane'. In either case, parallax errors occur under depth-varying scene conditions.

In this paper, we show a novel PoG estimation approach for remote and head-mounted systems that overcomes the issues using a *direct mapping* strategy (Fig. 1). Since the cornea of the human eye reflects the light from the surrounding

environment, an image of the eye contains a reflected image of the scene. Using this observation, we calculate the pose of the eye and the corneal surface location, where light from the PoG reflects into the eye image. Directly matching at the reflected PoG, then, obtains the location of the PoG in an environment image. This approach solves the discussed issues and allows for dynamic depth-varying environments as it does not require any mapping functions.

Therefore, this work introduces two novel points: First, we find the so-called gaze-reflection point (GRP), where light from the PoG reflects at the corneal surface into the eye camera. This relates to the forward projection problem, for which Agrawal et al. recently derived a first analytic solution [4]. Unfortunately, their method is not applicable for this more general case, where the PoG is defined up to an unknown distance along the gaze ray. We, therefore, introduce a novel analytic method for obtaining the GRP based on a geometric eye model and a reflection model of the corneal surface. We further provide a one-point calibration technique that is required once per user, and largely increases accuracy by compensating for the individual offset between optical and visual axis in human vision. Second, we introduce the idea of using active coded illumination for robust and accurate correspondence matching between a corneal image and a direct view of the scene (Fig. 1(d)). Because corneal reflections suffer from low image quality, low intensity, superimposed iris texture, and unknown geometric distortion, this kind of matching is difficult to achieve with passively captured images. To compensate for this, we describe an active strategy using a novel LED-array projector to assign scene locations with temporal-coded patterns that can be identified from within both, corneal and scene images.

The remainder of the paper is organized as follows: Section 2 surveys related work. Section 3 describes a geometric eye model and a reflection model for the corneal surface to develop a method for obtaining the GRP. Section 4 describes the concept of direct matching and introduces the coded illumination projector. Sections 5 to 7 show experimental results, followed by discussion and conclusion.

2 Related Work

In the following, we discuss the relation of this work to the fields of corneal reflection analysis, eye gaze tracking and active coded illumination.

Corneal Reflection Analysis: The environmental light distribution around a person can be obtained from corneal images. Backes et al. [5] estimate the point spread function of the corneal reflection system and apply it to recover clear environment images using non-blind deconvolution. Nishino and Nayar [6] formalize the imaging characteristics of the eye-camera geometry as a non-central catadioptric system [7] that requires per-frame calibration through eye pose estimation. They describe a number of applications including environment map extraction, gaze target (PoG) estimation and face relighting. However, the gaze target estimation technique has several drawbacks compared to our approach: (1) It requires a more complex eye model with known eyeball size, and the back

projection of the whole corneal image. (2) It does not model the individual offset between optical and visual axis, which is not accurate. (3) It obtains the PoG only in the (low quality) corneal image. (4) It uses passively captured corneal reflections that do not allow for robust matching.

Eye Gaze Tracking (EGT): Traditional PCCR approaches rely on a trained mapping from eye-image derived data (e.g., appearance, 2D or 3D geometric features) to the PoG on a target surface [2]. This approach is simple but involves a number of drawbacks. In particular, it assumes a static geometric relation between eye, camera, and a planar scene, which requires tedious calibration, and leads to decreased accuracy from unavoidable drift of the eye camera w.r.t. the head and a parallax variation with scene depth. More recent approaches achieve head-pose invariance by explicitly modeling the 3D geometry of the setup. The eye is tracked through a remote stationary high-resolution camera or a dynamic camera system [8,9]. In these systems, the PoG is obtained as the intersection of gaze ray and target surface. These approaches achieve reduced intrusiveness, but are more complex as they require domain knowledge. Also, they still involve a combination of drawbacks such as complicated extrinsic calibration of cameras and light sources, and accuracy degradation when the geometric relation changes. Several interesting techniques use machine learning to automatically obtain a gaze mapping to a computer screen by exploiting mouse click positions [10] or visual saliency maps of display content [11]. While not requiring a dedicated calibration or attached devices, this assumes particular application scenarios, relies on the properties of human perception, and obtains only low accuracy.

Coded structured light is widely used to identify environment locations, for scene reconstruction [12] and object tagging [13]. Non-intrusiveness for human observers is accomplished by exploiting the imperceptibility of complement patterns at high framerates or invisible IR light. Because commodity projectors are optimized for the visible spectrum, IR projection commonly requires special devices. Lee et al. [14] describe a high-resolution, scalable, and general-purpose solution that replaces the lamp of a DLP projector by an array of IR and visible-light LEDs; however, its luminance is insufficient for our application. Special LED arrays have been previously applied in EGT to allow automatic and robust extraction of corneal reflections [15]. While referred to as *structured light*, this relates only to the geometric alignment of the LED array.

3 Calculation of the Gaze Reflection Point

The GRP is obtained as follows: First, the three-dimensional pose of the eye w.r.t. the eye camera is found through a geometric model-based iris contour fitting. Then, the GRP is calculated using the eye pose and a reflection model at the corneal surface. In addition, a corrected GRP is obtained by compensating for the individual offset between optical and visual axis. We will now describe each step in detail.

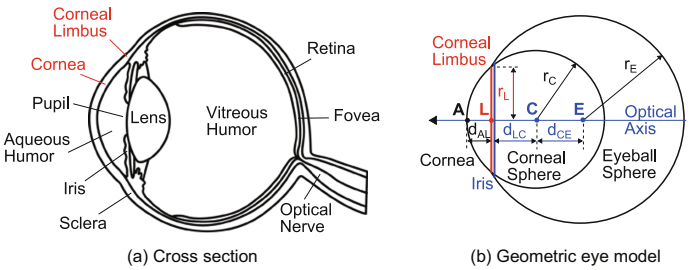


Fig. 2. (a) Cross section of the human eye and (b) geometric eye model with spherical curvature for eyeball and cornea

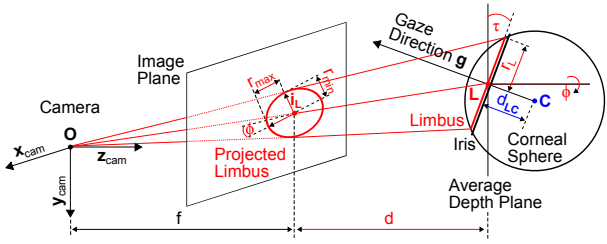


Fig. 3. 3D eye pose estimation from the projected limbus (iris contour)

3.1 Geometric Eye Model

Fig. 2(a) shows a cross-section of the eye. The human eyeball is not a plain sphere; it can be subdivided into two main segments: the anterior and the posterior. Both segments can be approximated by two overlapping spheres of different size (Fig. 2(b)). We use this geometric model to estimate the 3D pose of the eye and the location of the GRP in the image. Although there have been several discussions as to whether the eyeball shape is more complex than the model, and whether or not its anatomic parameters vary among humans, several previous studies have shown that this simplified model is sufficiently accurate [16]. The cornea is modeled as a spherical cap that is cut off from the corneal sphere by the limbus plane. The average radius of the limbus r_L is approximately 5.6 mm [17].

3.2 Eye Pose Estimation

We assume weak perspective projection since the depth of a tilted limbus is much smaller than the distance between eye and camera. Thus, the almost circular limbus projects to an ellipse that is described by five parameters: the center $\mathbf{i_L}$, the major and minor radii $r_{\max}$ and $r_{\min}$, and the rotation angle ϕ (Fig. 3). As the iris is delimited by the limbus, we estimate these values by fitting an ellipse to the projected iris contour.

Eye pose estimation is initialized from the pupil contour obtained through the dark-pupil method that exploits the relatively high light absorption of the

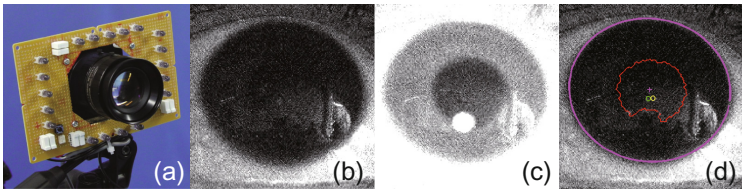


Fig. 4. Iris contour detection using the dark-pupil method. (a) The eye camera with a concentric ring of IR LEDs. (b) First frame (under normal lighting). (c) Second frame (under IR illumination). The bright spot denotes the reflection of the IR illumination. (d) Detected pupil contour (red) from the dark-pupil method. The magenta ellipse and '+' denote the estimated limbus contour and center position. The cyan circle indicates the center of the corneal sphere that coincides with the reflection of the IR illumination. Green and yellow circles indicate the GRP and the corrected GRP (GRP*).

pupil compared to other eye regions. The method uses two corneal images, taken under normal lighting and under illumination from a concentric ring of IR-LEDs around the camera lens (Fig. 4(a)–(c)). We subtract and threshold these frames to segment the pupil region, and apply a RANSAC-based ellipse-fitting to estimate its contour $B\left(x, y, c_x, c_y, a, b, \phi\right)$, where (c_x, c_y) is the center, a and b are the two radii, and ϕ is the rotation angle. We initialize the iris contour parameters using c_x, c_y, ϕ, and calculate the radii a_0, b_0 as in

$$eval_0\left(c_x, c_y, a, b, \phi\right) = \sum_{x,y} E_x\left(x, y\right) \cdot \operatorname{sgn}\left(c_x - x\right) \cdot B\left(x, y, c_x, c_y, a, b, \phi\right), \quad (1)$$

$$\left[a_0, b_0\right] = \arg\max_{a,b} eval_0\left(c_x, c_y, a, b, \phi\right), \quad (2)$$

where $E_x\left(x, y\right)$ are the x-differentials of the eye image and $a_0 > a$, $b_0 > b$. We only use x-differentials to avoid the effects of eyelid occlusion edges.

The gaze direction $\mathbf{g}$ is approximated by the optical axis of the eye, and theoretically obtained as in

$$\mathbf{g} = \left[\sin\tau \sin\phi \; -\sin\tau \cos\phi \; -\cos\tau\right]^{\mathrm{T}}. \quad (3)$$

The angle ϕ is known as the rotation of the limbus ellipse in the image plane. The angle $\tau = \pm \arccos\left(r_{\min}/r_{\max}\right)$ corresponds to the tilt of the limbus plane w.r.t. the image plane. The center of the corneal sphere $\mathbf{C}$ is located at distance $d_{\mathbf{LC}}(= 5.6$ mm) from the limbus center[1] $\mathbf{L}$ along $-\mathbf{g}$. The center $\mathbf{C}$ and the radius r_C, approximately 7.7 mm [17], describe the corneal sphere that enables to model the light-reflection properties of the corneal surface.

Finally, we apply an iterative optimization to obtain accurate eye pose parameters. Since the cornea is a spherical surface, its projected center must coincide

[1] The position of the limbus center can be calculated from the center of the detected ellipse and the distance to the camera $d = f \cdot r_L / r_{\max}$ as in $\mathbf{L} = d\mathbf{K}^{-1}\mathbf{i}_{\mathbf{L}}$, where $\mathbf{K}$ is the camera matrix and f the focal length in pixels. However, this is not required for the calculation of the GRP.

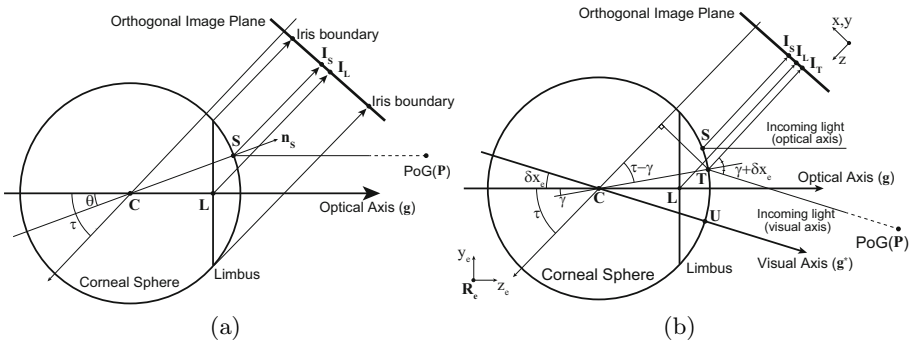


Fig. 5. (a) Corneal reflection and gaze-reflection point (GRP). The light-ray parallel to the gaze direction reflects at the GRP into the image of an eye-observation camera. (b) The optical axis and visual axis of the human eye. The light ray from the optical axis is reflected at the GRP **S**, and the calibration point is reflected at the calibration reflection point (CRP) **T**. The angular difference between these two axes (δx, δy) is obtained to compensate the individual offset. This illustration shows the cross section at the short axis of the projected limbus.

with the the centroid of the IR-LED ring reflection that represents the corneal image of the camera itself (Fig. 4(a),(c)). We formulate this constraint as an additional error term in equation (1) as in

$$eval_1\left(c_x, c_y, a, b, \phi\right) = \sum_{x,y} E_x\left(x, y\right) \cdot \operatorname{sgn}\left(c_x - x\right) \cdot B\left(x, y, c_x, c_y, a, b, \phi\right) + \alpha \cdot \left\|\mathbf{p}_{\mathrm{LED}} - \mathbf{p}_{\mathrm{CC}}\right\|, \tag{4}$$

and estimate the optimal iris contour parameters through iterative minimization. Here, $\mathbf{p}_{\mathrm{LED}}$ is the surface reflection point of the IR-LED, $\mathbf{p}_{\mathrm{CC}}$ is the projected corneal center, and α is a positive constant that is determined experimentally. The combined image- and geometric-model-based evaluation enables increased accuracy and robustness compared to existing iris-contour based eye pose estimation methods.

3.3 GRP Estimation from the Optical Axis of the Eye

According to the law of reflection, possible solutions for the GRP are not arbitrarily distributed on the surface of the corneal sphere, but on a circular subspace, arising from the intersection with the plane of reflection, that contains incident ray, reflected ray, and surface normal $\mathbf{n_S}$, and, thus, camera center **O**, corneal sphere center **C**, PoG **P**, and GRP **S**. Assuming weak perspective projection, the plane of reflection intersects the semi-minor axis of the projected limbus (ellipse) and is orthogonal to the image plane. Fig. 5(a) illustrates the

light reflection in the plane. We can formulate the reflection of the light ray parallel to the optical axis of the eye at the GRP as follows:

$$\mathbf{C}\cdot\mathbf{n_S} = \mathbf{g}\cdot\mathbf{n_S}, \quad \mathbf{C} = \begin{bmatrix}0\ 1\end{bmatrix}^{\mathrm{T}}, \quad \mathbf{g} = \begin{bmatrix}\cos\tau\ \sin\tau\end{bmatrix}^{\mathrm{T}}, \quad \mathbf{n_S} = \begin{bmatrix}\cos\theta\ \sin\theta\end{bmatrix}^{\mathrm{T}}. \tag{5}$$

We then solve for the corneal angle θ that defines the GRP, using the known eye gaze angle τ as in

$$\begin{aligned}\theta &= \arctan((1-\sin\tau)/\cos\tau)\\ &= \tau/2.\end{aligned} \tag{6}$$

The distance between the limbus center and the GRP in the local orthogonal image plane is defined as in

$$|\mathbf{I_S} - \mathbf{I_L}| = r_C \sin\theta - d_{\mathbf{LC}} \sin\tau. \tag{7}$$

Assuming the average depth plane for weak perspective projection to be located at the center of the limbus $\mathbf{L}$, then, the location of the GRP $\mathbf{i_{S'}}$ is given as in

$$\mathbf{i_S} = \mathbf{i_L} + s \cdot \mathbf{v}_{sm} \left|\mathbf{I_S} - \mathbf{I_L}\right|, \quad s = r_{\max}/r_L, \tag{8}$$

where s is the scale factor of the weak perspective projection, $\mathbf{i_L}$ is the center of the limbus ellipse, and $\mathbf{v}_{sm}$ is its semi-minor axis.

3.4 Calibration and Application of Individual Parameters

Many studies report that the optical axis of a human eye differs by an individual offset from the visual axis, which is the true gaze direction [2]. This difference can be described by static offset angles in a facial coordinate system, where z_{face} is the facial frontal direction, x_{face} is the direction from left to right eye, and y_{face} is orthogonal to x_{face} and z_{face}. The offset angles around x_{face} (tilt) and y_{face} (pan) are commonly within $1.5 \sim 3$ and $4 \sim 5$ deg. However, a one-time personal calibration is needed to obtain accurate values for a particular user. We introduce a one-point calibration technique to obtain these parameters through the analysis of a corneal reflection image.

In the calibration process, a user gazes a calibration point, such as a highly illuminated scene location or marker. We calculate the GRP and manually find the reflection of the calibration point (calibration reflection point (CRP)) in the corneal reflection image. We then introduce a coordinate system $\mathbf{R}_{\mathrm{e}} = [\mathbf{x}_{\mathrm{e}}\, \mathbf{y}_{\mathrm{e}}\, \mathbf{z}_{\mathrm{e}}]$ at the limbus center, where $\mathbf{x}_{\mathrm{e}}$ and $\mathbf{y}_{\mathrm{e}}$ correspond to the three-dimensional vectors of the long and short axes of the projected limbus, respectively. Fig. 5(b) illustrates the cross section at the short axis ($\mathbf{y}_{\mathrm{e}}\mathbf{z}_{\mathrm{e}}$-plane). In this plane, the angle γ between the optical axis and the CRP is described as in

$$|\mathbf{I_L} - \mathbf{I_T}| = \frac{|\mathbf{v}_{sm} \cdot (\mathbf{i_T} - \mathbf{i_L})|}{s}, \tag{9}$$

$$\gamma = \tau - \arcsin(d_{\mathbf{LC}} \cdot \cos\tau + \frac{|\mathbf{I_L} - \mathbf{I_T}|}{r_C}), \tag{10}$$

where $\mathbf{I_T}$ and $\mathbf{i_T}$ are the location of the CRP and its projection in the corneal image, respectively. Considering a surface reflection at the CRP, then the offset angle δx_{e} between the optical and the visual axis in this plane is defined as in

$$2 \cdot (\gamma + \delta x_{\mathrm{e}}) = \tau + \delta x_{\mathrm{e}}, \tag{11}$$

$$\delta x_{\mathrm{e}} = \tau - 2 \cdot \gamma. \tag{12}$$

Similarly, the angle δy_e in the $\mathbf{x}_{\mathrm{e}}\mathbf{z}_{\mathrm{e}}$-plane is defined as in

$$\delta y_e = 2 \cdot \arcsin(\frac{\mathbf{v}_{lm} \cdot (\mathbf{i_T} - \mathbf{i_L})}{s \cdot r_C}), \tag{13}$$

where $\mathbf{v}_{lm}$ is the major axis of the limbus ellipse.

Assuming the camera to be located in front of the face, the offset angles can be transformed into the camera coordinate system as in

$$\delta x = \arctan(\frac{\mathbf{e}_{\mathrm{y}}^{\mathrm{T}} \mathbf{R}_{\mathrm{e}} \mathbf{u}}{\mathbf{e}_{\mathrm{z}}^{\mathrm{T}} \mathbf{R}_{\mathrm{e}} \mathbf{u}}), \tag{14}$$

$$\delta y = \arctan(\frac{\mathbf{e}_{\mathrm{x}}^{\mathrm{T}} \mathbf{R}_{\mathrm{e}} \mathbf{u}}{\mathbf{e}_{\mathrm{z}}^{\mathrm{T}} \mathbf{R}_{\mathrm{e}} \mathbf{u}}), \tag{15}$$

$$\mathbf{u} = \left[\sin \delta y_{\mathrm{e}} \sin \delta x_{\mathrm{e}} \sqrt{1 - \sin^2 \delta x_{\mathrm{e}} - \sin^2 \delta y_{\mathrm{e}}}\right]^{\mathrm{T}}, \tag{16}$$

$$\mathbf{e}_{\mathrm{x}} = [1\ 0\ 0]^{\mathrm{T}},\ \mathbf{e}_{\mathrm{y}} = [0\ 1\ 0]^{\mathrm{T}}, \mathbf{e}_{\mathrm{z}} = [0\ 0\ 1]^{\mathrm{T}}.$$

Once the transformed offset angles are obtained, the corrected GRP (GRP*) can be estimated using the corrected eye coordinate system $\mathbf{R}_{\mathrm{e}}^*$ that is rotated by $(\delta x, \delta y)$ from the original eye coordinate system $\mathbf{R}_{\mathrm{e}}$. Using $\mathbf{R}_{\mathrm{e}}^*$, we obtain the corrected angle τ^* and the semi-minor axis of the projected limbus $\mathbf{v}_{sm}^*$ as in

$$\tau^* = \arccos(\mathbf{e}_{\mathrm{z}}^{\mathrm{T}} \mathbf{R}_{\mathrm{e}}^* \mathbf{e}_{\mathrm{z}}), \tag{17}$$

$$\mathbf{v}_{sm}^* = (\begin{bmatrix} \mathbf{e}_{\mathrm{x}}^{\mathrm{T}} \\ \mathbf{e}_{\mathrm{y}}^{\mathrm{T}} \end{bmatrix} \mathbf{R}_e^* \mathbf{e}_{\mathrm{y}}) / \left\| \begin{bmatrix} \mathbf{e}_{\mathrm{x}}^{\mathrm{T}} \\ \mathbf{e}_{\mathrm{y}}^{\mathrm{T}} \end{bmatrix} \mathbf{R}_e^* \mathbf{e}_{\mathrm{y}} \right\|, \tag{18}$$

$$\mathbf{R}_{\mathrm{e}}^* = \mathbf{R}_x(\delta x) \mathbf{R}_y(\delta y) \mathbf{R}_{\mathrm{e}}, \tag{19}$$

where $\mathbf{R}_x$ and $\mathbf{R}_y$ are the rotation matrices around the x and y axes, respectively. Applying τ^*, $\mathbf{v}_{sm}^*$, and equations (7) and (8), we obtain the position of the GRP* that reflects the light from the true visual axis of the particular eye.

4 Direct Matching Using Coded Illumination

4.1 LED Array Projector

Coded illumination is the key component for matching the GRP in a corneal reflection image to find the PoG in the scene. For this purpose, we designed a novel programmable LED-array projector (LED-AP) that produces high-power and high-frequency coded light beams to achieve real-time PoG estimation (Fig. 6).

It comprises an array of LED-lens units, LED power modules and a controller (Fig. 7). We evaluate two types of LEDs: high-power visible-light and IR LEDs to produce imperceptible IR illumination. Each LED can be toggled on/off within 0.05 ms according to control signals from an Arduino micro-controller. As each LED's illumination is modulated with a unique temporal coding sequence, the system can identify projected light-spots by decoding the series of eye (Fig. 8(c)–(e)) and environment images. The direction of an LED can be manually adjusted so that the illumination covers the whole scene or particular objects of interest. The number of LEDs defines the spatial resolution of the coverage area, which is usually lower than that of data projectors. To compensate for this, we employ the following linear interpolation technique for PoG estimation: First, we recover the codes from the sequences of corneal reflection and environment images, and obtain the center of the moments of the LED light-spots. In the corneal reflection image, we calculate the relative position of the GRP w.r.t. four neighboring light-spots, and use this to locate the PoG in the environment image.

4.2 Coding Strategy

For temporal modulation of the LED-AP illumination, we implemented both, a binary code and a Hamming(7,4) code [18]. While the binary code is simple and short-length, the Hamming(7,4) code allows to correct a single-bit errors but encodes 4 data bits into 7 bits by adding 3 parity bits. Each implementation requires one frame for pupil detection using the dark-pupil method before starting a temporal code block. Fig. 9 shows the signal ordering.

5 Experiments

We evaluate the proposed method and compare it to a mapping-based PCCR technique that is commonly used in existing EGT systems.

PCCR is implemented using two kinds of IR illumination: a previously described on-axis IR-LED ring around the camera lens to realize the dark-pupil method and a single off-axis IR-LED that is static w.r.t. the eye camera. In the calibration step, four eye images are captured, where a subject gazes the corners of a rectangle on a calibration plane. Detecting pupil centers and glints from the off-axis IR-LED allows to calculate an affine matrix that maps the pupil center/glint vector to the corresponding calibration point in the scene coordinate system. At the gaze estimation step, this mapping function is used to obtain the PoG from the pupil center and glint in a corneal image.

Experimental Setup: We performed experiments with five subjects in an indoor environment under normal lighting conditions. Fig. 7 (d) shows the experimental setup. We use Point Grey Dragonfly Express cameras (640×480 pixels, B/W, 30 Hz) and MATLAB 2011b with the Image Processing Toolbox on a PC (Linux, Intel Core i7-960 3.06 GHz, 8 GB RAM). The eye camera has a horizontal angle of view of approximately 9 deg, where the projected limbus radius is approximately 100 pixels. To evaluate the effects of depth variation in

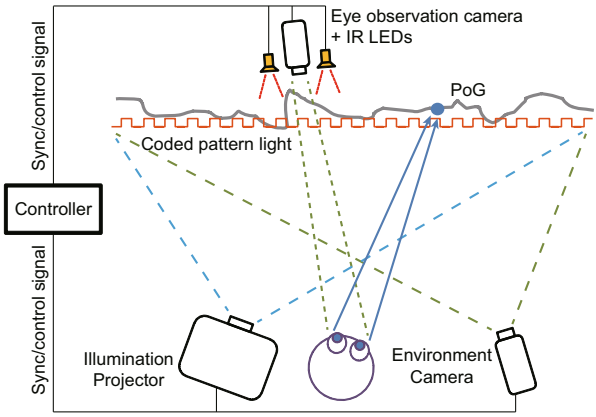


Fig. 6. System configuration. The system consists of an eye-observation camera with IR LEDs, an environment camera and an LED-array projector.

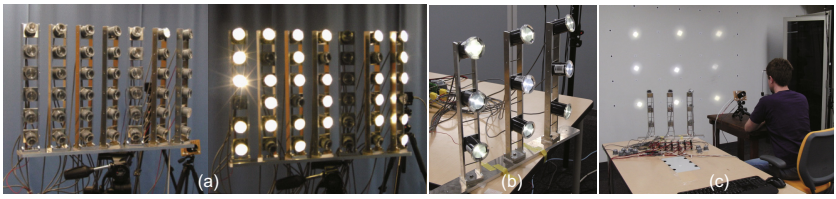


Fig. 7. LED array projector (LED-AP). (a) LED array (42-LED configuration). Each unit consists of a high-luminance LED (visible light or IR-light) with a Fresnel lens. (b) LED array (9-LED configuration). (c) Experimental setup and projected illumination patterns.

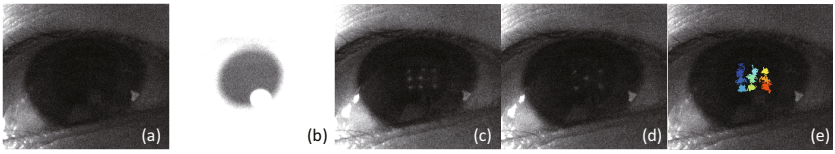


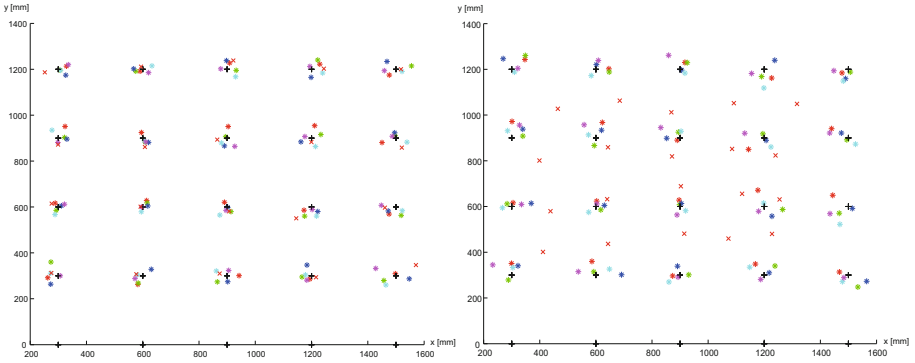
Fig. 8. Corneal images obtained with the LED-AP implementation. (a) Frame #1 (under normal lighting). (b) Frame #2 (under IR illumination). (c) Frame #3, (d) frame #13, and (e) decoded result from a sequence of temporal-coded patterns with the reflections of 9 light spots.

Frame #	1	2	3	4	5	....
IR LED RING	OFF	ON	OFF	OFF	OFF	....
LED-AP	OFF	OFF	ALL ON	CODE#1	CODE#2	....

Fig. 9. Signal ordering for the concentric ring of IR-LEDs attached around the lens of the eye camera, and the LED-AP. At frames #1 and #2, the IR-LED ring is turned on and off for pupil detection. At frame #3, the LED-AP is completely turned on to obtain the reference illumination for subsequent signal code frames.

Table 1. Mean PoG estimation error [deg]

	Condition 1 (depth = 1900 mm)		Condition 2 (depth = 3000 mm)	
	Proposed method	PCCR	Proposed method	PCCR
User 1	0.960	0.973	0.909	3.084
User 2	0.913	1.032	0.642	4.350
User 3	0.916	0.963	0.756	3.560
User 4	0.947	0.890	0.759	3.951
User 5	0.761	1.121	0.914	4.240
Mean	**0.899**	0.996	**0.796**	3.837



(a) Condition 1 (depth = 1900 mm) (b) Condition 2 (depth = 3000 mm)

Fig. 10. Result of PoG estimation under condition 1 and 2. '+' indicate ground-truth markers, and '∗' indicate PoG estimation results from the proposed method. 5 users are represented by different colors. Red 'x' indicate results of implicit EGT for user 1. The light-spots of the 9 LEDs are located at the plane coordinates (1490, 1200), (895, 1200), (300, 1200), (1430, 750), (835, 750), (300, 750), (1450, 300), (820, 300) and (320, 300) [mm]. The eye positions are located at (900, 600, 1900) and (900, 600, 3000) [mm] at condition 1 and 2, respectively.

the scene, subjects are seated at both, 1900 (condition 1) and 3000 mm (condition 2), in front of a wall, where 20 ground-truth markers are attached. The subjects are asked to gaze the markers, where we calculate the angular errors between ground-truth and estimated PoGs. Individual parameters are calibrated at condition 1 and also applied at condition 2. The eye camera is placed slightly below a subject's face at a distance of 600 mm in front of the eyes. The environment camera is placed behind the subject so as to capture the entire scene. We used 9-LED illumination for these experiments, where the coordinates of the light-spots are described in Fig. 10.

Results: Table 1 shows the angular errors, obtained with the subject's right eyes are located at $(x, y, z) = (900, 600, 1900)$ and $(900, 600, 3000)$ [mm]. Fig. 10 shows the PoG estimation results at conditions 1 and 2. The proposed

method results in an error of less than 1 deg at both conditions. However, the PCCR method suffers from a reduced performance at condition 2. Please refer to a supplementary video[2] for results in more complex environments.

6 Discussion

The proposed method performs with considerable accuracy, although it does not require body-attached devices and allows free head movement. A novel one-point personal calibration technique successfully increased the accuracy under all experimental conditions. Moreover, we confirmed that the method supports PoG estimation under a varying scene depth.

The performance of the implemented single-glint PCCR technique is found to decrease when the depth of the scene changes from the one at which calibration was performed, which relates to an increasing parallax error for the calibrated mapping function. On the contrary, the proposed approach relies on an uncalibrated direct matching between corneal reflection and environment image, which performs accurately and stably under varying depth conditions.

We implemented and evaluated two temporal-coding patterns for LED-AP illumination. Even though, the Hamming(7,4) code requires more frames than the binary code, there was only little difference in the decoding results for corneal reflection images under the current experimental setup.

Eyelid occlusion is a common problem in any EGT system. In the proposed approach, this does not affect the PoG mapping as the GRP is located near the center of the pupil. However, it increases the error in iris-contour based eye pose estimation, where more robust and accurate strategies need to be investigated.

PoG estimation using the LED-AP assumes local planarity between the light spots in the scene. Increasing the intensity and number of LEDs can, therefore, potentially increase the accuracy under complex geometry and decoding errors, e.g., from object occlusion and noise.

7 Conclusion

The paper introduced a novel point of gaze estimation scheme that overcomes two major limitations of existing methods, namely, (1) the requirement for calibration of a static relationship between eye, eye camera and scene, and (2) a parallax error that occurs when the depth of the scene varies. The proposed approach comprises two major contributions: First, an analytic solution is developed to calculate the gaze reflection point (GRP), where light from the point of gaze (PoG) reflects into an eye image. This includes compensation for the individual offset between optical and visual axis. Second, active coded illumination enables robust and accurate matching at the GRP to obtain the PoG in a scene image, which is also the first use of this technique in EGT and corneal reflection analysis. Experimental evaluation with a specially-designed high-power IR LED-array projector shows that the proposed scheme achieves considerable accuracy and successfully supports depth-varying environments.

[2] `http://www.ime.cmc.osaka-u.ac.jp/~nakazawa/wiki/index.php?EGT`

Acknowledgement. This work was supported by the JST Precursory Research for Embryonic Science and Technology (PRESTO) program, Japan.

References

1. Young, L., Sheena, D.: Survey of eye movement recording methods. Behav. Res. Meth. Instrum. 7, 397–429 (1975)
2. Hansen, D.W., Ji, Q.: In the eye of the beholder: A survey of models for eyes and gaze. IEEE Trans. Pattern Anal. Mach. Intell. 32, 478–500 (2010)
3. Guestrin, E.D., Eizenman, M.: General theory of remote gaze estimation using the pupil center and corneal reflections. IEEE Trans. Biomed. Eng. 53, 1124–1133 (2006)
4. Agrawal, A., Taguchi, Y., Ramalingam, S.: Beyond Alhazen's problem: Analytical projection model for non-central catadioptric cameras with quadric mirrors. In: Proc. CVPR, pp. 2993–3000 (2011)
5. Backes, M., Chen, T., Dürmuth, M., Lensch, H.P.A., Welk, M.: Tempest in a teapot: Compromising reflections revisited. In: Proc. IEEE SP, pp. 315–327 (2009)
6. Nishino, K., Nayar, S.K.: Corneal imaging system: Environment from eyes. Int. J. Comput. Vision 70, 23–40 (2006)
7. Sturm, P., Ramalingam, S., Tardif, J.P., Gasparini, S., Barreto, J.: Camera models and fundamental concepts used in geometric computer vision. Found. Trends. Comput. Graph. Vis. 6, 1–183 (2011)
8. Kim, S.M., Sked, M., Ji, Q.: Non-intrusive eye gaze tracking under natural head movements. In: Proc. IEEE IEMBS, pp. 2271–2274 (2004)
9. Reale, M., Hung, T., Yin, L.: Viewing direction estimation based on 3D eyeball construction for HRI. In: Proc. CVPRW, pp. 24–31 (2010)
10. Sugano, Y., Matsushita, Y., Sato, Y., Koike, H.: An Incremental Learning Method for Unconstrained Gaze Estimation. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 656–667. Springer, Heidelberg (2008)
11. Sugano, Y., Matsushita, Y., Sato, Y.: Appearance-based gaze estimation using visual saliency. IEEE Trans. Pattern Anal. Mach. Intell., 1 (2012)
12. Salvi, J., Fernandez, S., Pribanic, T., Llado, X.: A state of the art in structured light patterns for surface profilometry. Pattern Recogn. 43, 2666–2680 (2010)
13. Zhang, L., Subramaniam, N., Lin, R., Raskar, R., Nayar, S.: Capturing images with sparse informational pixels using projected 3D tags. In: Proc. IEEE VR, pp. 11–18 (2008)
14. Lee, J.C., Hudson, S., Dietz, P.: Hybrid infrared and visible light projection for location tracking. In: Proc. ACM UIST, pp. 57–60 (2007)
15. Li, F., Kolakowski, S., Pelz, J.: Using structured illumination to enhance video-based eye tracking. In: Proc. IEEE ICIP, pp. 373–376 (2007)
16. Nitschke, C., Nakazawa, A., Takemura, H.: Display-camera calibration using eye reflections and geometry constraints. Comput. Vis. Image Underst. 115, 835–853 (2011)
17. Snell, R.S., Lemp, M.A.: Clinical Anatomy of the Eye, 2nd edn. Blackwell Publishing, Malden (1997)
18. Moon, T.K.: Error Correction Coding: Mathematical Methods and Algorithms. Wiley-Interscience (2005)