GAZE TRACKING USING CORNEAL IMAGES CAPTURED BY A SINGLE HIGH-SENSITIVITY CAMERA

L. El Hafi, M. Ding, J. Takamatsu, and T. Ogasawara

Robotics Laboratory, Graduate School of Information Science,
Nara Institute of Science and Technology (NAIST), Japan

ABSTRACT

This paper introduces a method to estimate gaze direction using images of the eye captured by a single high-sensitivity camera. The purpose is to develop wearable devices that enable intuitive eye-based interactions and applications. Indeed, camera-based solutions, as opposed to commercially available infrared-based ones, allow wearable devices to not only obtain natural user responses from eye movements, but also scene images reflected on the cornea, without the need for additional sensors. The proposed method relies on a model approach to evaluate the gaze direction and does not require a frontal camera to capture scene information, making it more socially acceptable if embedded in a glasses-shaped device. Moreover, recent development in high-sensitivity camera sensors allows us to consider the proposed method even in low-light condition. Finally, experimental results using a prototype wearable device demonstrate the potential of the proposed method solely based on cornea images captured from a single camera.

INTRODUCTION

Today's high-resolution, high-sensitivity cameras alongside powerful image processing algorithms make possible many new applications. In particular, the increase in resolution and the deceased size of camera sensors allow new eye-tracking methods previously judged impractical.

Moreover, the industry's recent growing interest in virtual reality (VR), augmented reality (AR) and smart wearable devices has created a new momentum for eye tracking. Indeed, eye tracking can be used as an intuitive AR input, or used to reduce motion sickness induced by ill-calibrated VR devices (1). Eye movements in particular are viewed as a way to obtain natural user responses from wearable devices alongside gaze information to analyze interests and behaviors (2).

In this paper, we introduce a method to estimate the gaze direction using cornea images captured by a single high-sensitivity camera. Corneal imaging was first explored in (3) and further refined in (4), (5) and (6). Camera-based solutions, as opposed to commercially available infrared-based (IR) ones, allow wearable devices not only to obtain natural user responses from eye movements, but also scene images reflected on the cornea without the need for additional sensors. In particular, our method does not require a frontal camera to capture the scene, making it more socially acceptable as part of a wearable device.

We use a model-based approach to estimate the gaze direction in our proposed method. First, we reconstruct a 3D eye model from an image of the eye by fitting an ellipse on the colored iris area. Then we continuously track the gaze direction by rotating the model to simulate projections of the iris area for different eye poses and matching the iris area of the subsequent images with the corresponding projections obtained from the model. From an additional one-time calibration step, we can also compute the reflected point of regard on the cornea, enabling us to identify where a user is looking in the scene image reflected on the cornea.

In order to validate our method, we conducted several experiments using different hardware, such as a high-sensitivity camera in low-light condition and glasses equipped with a near-4K camera. We did this in front of a computer display to demonstrate the potential of such an eye-tracking method based solely on cornea images captured from a single camera.

The remainder of this paper is structured as follows. First, we briefly introduce a geometric model derived from the main characteristics of the human eye. Second, we describe how to build a 3D eye model from an image of the eye and estimate both its location and orientation relative to the camera. Third, we propose a method to continuously track the gaze direction using the previously built model. Fourth, we present the experimental results obtained using a high-sensitivity camera in low-light condition as well as prototype glasses. Fifth and finally, we conclude by suggesting further areas of work to investigate.

EYE MODELIZATION

This section describes the main characteristics of the human eye and how to derive a geometric model from them.

Human Eye

Figure 1 shows a cross-section of the human eye. When observing an eye from the outside, the most distinctive parts are the colored iris, the pupil at its center and the white sclera that surrounds it, as described in (4). The outer layer of the front of the eye is the cornea, which is more difficult to observe. It covers the iris and fades into the sclera at the limbus. The cornea and lens focus images onto the retina, or more precisely, onto the fovea which is the most sensitive part of the eye. Important properties of the cornea are its transparency and its specular reflection characteristics due to the film of tears that coats its surface. This mirror-like characteristic will be particularly relevant for extracting information about the scene and the point of regard (POR).

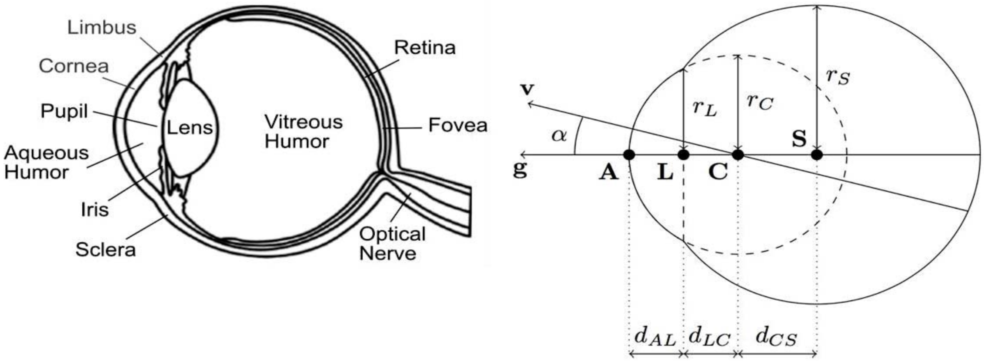

Figure 1 – Cross-section (4) and geometric model of the human eye

Eye Geometric Model

The human eye can be subdivided into two overlapping spheres of different sizes: a smaller sphere that includes the cornea, the iris, the pupil and the lens, and a bigger sclera sphere that includes the sclera, the vitreous humor and the retina with its fovea. The two spheres intersect at the limbus which defines a circle. This model is described in Figure 1:

- Points $\boldsymbol{L}$, $\boldsymbol{C}$ and $\boldsymbol{S}$ are respectively the limbus, cornea and sclera centers. A priori unknown.
- The vector $\boldsymbol{g}$ is the optical axis of the eye, crossing all the aforementioned centers. It intersects the cornea sphere at the cornea apex designated by $\boldsymbol{A}$.
- The vector $\boldsymbol{v}$ is the visual axis that goes from the fovea to the actual POR. The visual axis corresponds to the gaze direction and its estimation is the purpose of any gaze-tracking system.
- Distances r_L, r_C and r_S are respectively the limbus, cornea and sclera radii. Anatomical parameters.
- Distances d_{AL}, d_{LC} and d_{CS} separate the different components of the model. All are known anatomical parameters.

The optical axis is easy to estimate from the geometric properties of the eye but the visual axis is not. However, even though the visual axis, not the optical axis, corresponds to the direction of the POR, the optical axis can be used as a first approximation of the visual axis. The angle formed by the two axes is denoted by α and assumed to be constant.

This geometric model will be applied throughout this paper to estimate the pose of the eye from an image. It will also be used to describe interactions between the incident light and the cornea surface. By nature, such a model can only approximate the reality: the actual shape of the eye is more complex than the one described by the model and its anatomical parameters vary between individuals. However, we will assume that the variation of parameters among individuals is sufficiently small.

EYE POSE ESTIMATION

Now that the geometric model is defined, we build in this section a 3D model of the eye and estimate both its location and orientation relative to the camera.

Ellipse Fitting

We assume weak perspective projection since the depth of the tilted limbus is much smaller than the distance between the eye and the camera, as initially proposed in (4). Thus, the almost circular limbus projects to an ellipse described by five parameters: the center coordinates (c_u, c_v), the radii r_{max} and r_{min}, and the tilt ϕ as shown in Figure 2. Their values are estimated by fitting an ellipse on a set of limbus points, automatically detected or manually inputted, using least squares.

Pose Estimation

Now that the limbus has been fully described on the image plane, we can reconstruct a 3D model of the eye and estimate its pose in the world coordinate frame, i.e. estimate the coordinates of the limbus center $\boldsymbol{L}$ and the direction of the optical axis $\boldsymbol{g}$. The following geometric construction was originally proposed by (4). The origin $\boldsymbol{O} = (0, 0, 0)^{\mathrm{T}}$ is at the center of the camera lens as shown in Figure 2.

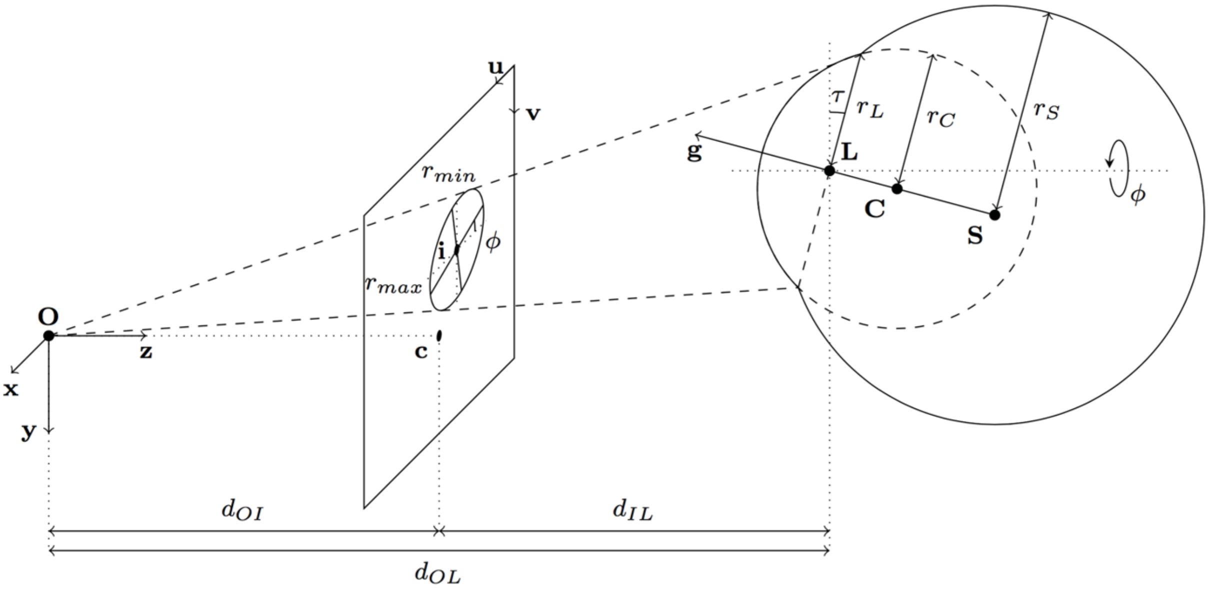

Figure 2 – 3D eye model construction

When the camera focus is assumed to be at infinity[1], d_{OL} can be expressed as:

$$d_{OL} = f\frac{r_L}{r_{max}}$$

where r_L, r_{max} and the focal length of the camera f are known.

If the limbus center is defined as $\boldsymbol{L} = \left(L_x,\ L_y,\ L_z\right)^{\mathrm{T}}$, we have by similarity:

$$\frac{L_x}{(i_u - c_u)s_x} = \frac{d_{OL}}{f},$$

$$\frac{L_y}{(i_v - c_v)s_y} = \frac{d_{OL}}{f},$$

where s_x and s_y are the pixel-to-world-unit scaling coefficients, obtained from camera calibration, respectively along the x and y directions.

By combining these equations, we have:

$$\boldsymbol{L} = \left(\frac{d_{OL}(i_u - c_u)s_x}{f}, \frac{d_{OL}(i_v - c_v)s_y}{f}, d_{OL}\right)^{\mathrm{T}}.$$

The tilt τ of the limbus plane with respect to the image plane is estimated from the shape of the ellipse up to a sign ambiguity:

---

[1] Which means $f = d_{OI}$. At close range, this sometimes cannot be assumed depending on the sizes of the sensor and the lens (6). To solve this problem, we can use a thin lens model:

$$\frac{1}{f} = \frac{1}{d_{OI}} + \frac{1}{d_{OL}},$$

where $f \neq d_{OI}$ and d_{OL} is required to compute d_{OI}. In the case of a head-mounted device, d_{OL} can be assumed to be known and constant.

$$\tau = \pm \arccos\left(\frac{r_{min}}{r_{max}}\right).$$

Indeed, two different limbus poses are possible from the projection alone: one looking in the direction of positive values of y, and another looking in the direction of negative values of y. In the case of a head-mounted camera, the ambiguity can be easily solved by knowing the relative pose of the camera to the eye, which is usually fixed and sufficiently tilted to avoid any ambiguity[2].

The optical axis $\boldsymbol{g}$ is then given by:

$$\boldsymbol{g} = (\sin\tau\sin\phi, -\sin\tau\cos\phi, -\cos\tau)^{\mathrm{T}},$$

where ϕ is already known as the rotation angle of the ellipse fitted on the limbus in the image plane.

Finally, the cornea center $\boldsymbol{C}$ and the sclera center $\boldsymbol{S}$ are given by:

$$\boldsymbol{C} = \boldsymbol{L} - d_{LC}\boldsymbol{g},$$

$$\boldsymbol{S} = \boldsymbol{L} - (d_{LC} + d_{LC})\boldsymbol{g},$$

and the limbus is computed as the intersection between the cornea and sclera spheres.

Visual Axis Calibration

To evaluate the direction of the visual axis $\boldsymbol{v}$, an additional calibration step is required. Figure 3 describes the relationship between the visual axis and the incident light coming from the POR, where $\boldsymbol{P}$ is the POR and $\boldsymbol{R}$ the reflected POR, i.e. the POR in the scene reflected on the cornea image. $\boldsymbol{n} = x\boldsymbol{s} + y\boldsymbol{l}$ is the normal at the reflected POR with $\boldsymbol{l}$ and $\boldsymbol{s}$ respectively the directions to the POR and to the camera optical center. Normal parameters x and y are unknown.

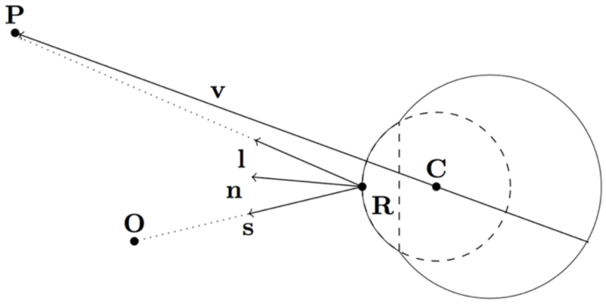

Figure 3 – Visual axis calibration

In the current implementation, the user is asked to manually register the reflected POR on the image. Assuming that the distance from the reflected POR to the POR is known during the calibration process, the direction of the visual axis can be computed using a specular model of a sphere (7).

---

[2] If not, we can also solve this ambiguity by attaching two LEDs to the camera (12). A line between the camera origin $\boldsymbol{O}$ and the cornea center $\boldsymbol{C}$ intersects at a point where the camera origin is reflected on the cornea. Therefore, we can obtain the projected cornea center location by finding the mean point of the two LEDs reflected in the image and use this information to resolve the sign ambiguity of τ.

The first step is to compute the normal $\boldsymbol{n}$ in order to find $\boldsymbol{R}$. This consists of solving the following biquadratic equation:

$$4cdy^4 - 4dy^3 + (a + 2b + c - 4ac)y^2 + 2(a - b)y + a - 1 = 0,$$

where $a = \boldsymbol{s} \cdot \boldsymbol{s}$, $b = \boldsymbol{s} \cdot \boldsymbol{l}$, $c = \boldsymbol{l} \cdot \boldsymbol{l}$, $d = \|\boldsymbol{s} \times \boldsymbol{l}\|^2$ are the coefficients. When $x = (2y^2 + y + 1) / (2by + 1)$ is defined, the normal $\boldsymbol{n}$ is computed from the solution in $x > 0$ and $y > 0$. The reflected POR $\boldsymbol{R}$ and the visual axis $\boldsymbol{v}$ are then computed by straightforward vector geometry.

GAZE TRACKING

In order to track the optical axis direction from the current image of the eye, we first attach a pitch-yaw-roll reference frame to the sclera center $\boldsymbol{S}$ of the 3D model built previously. The yaw axis is aligned with the eye corners for convenience. The pitch and yaw angles are respectively denoted by θ and ψ. By rotating the model around the pitch and yaw axis, we simulate several limbus projections for different eye poses, as shown in Figure 4. Note that we do not consider the roll angle assuming the human eye is not capable of such a rotation.

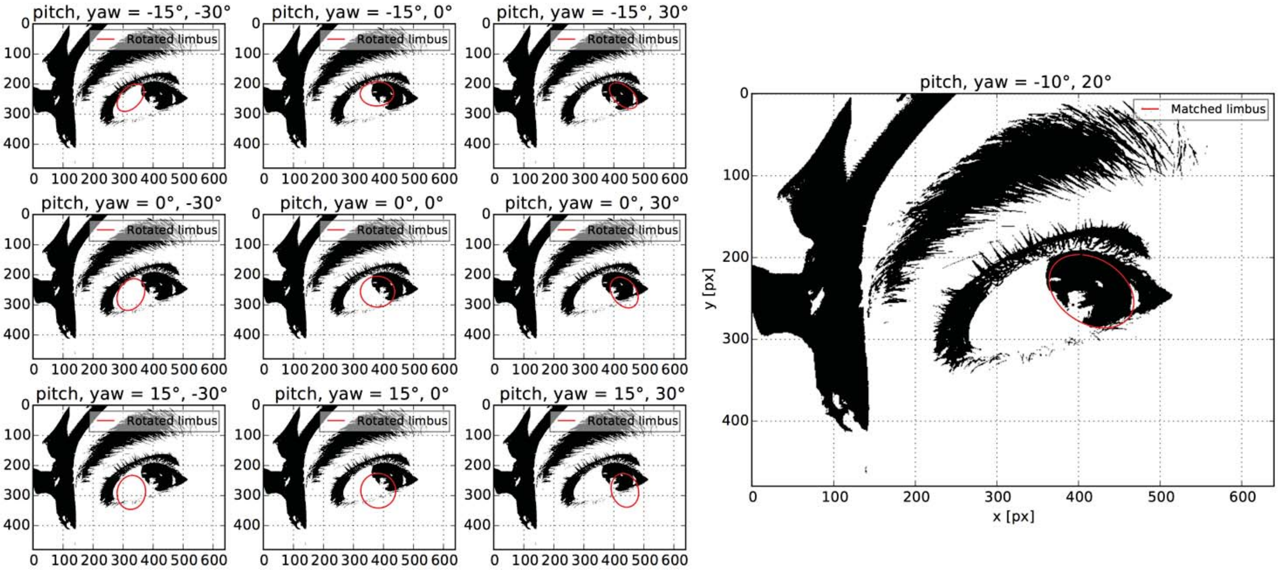

Figure 4 – Projection and matching of the rotated limbus into the image plane

The limbus of the rotated model is then projected into a binary image to serve as a mask. The projected area is set as binary 1s. The projection that matches the current limbus pose is then detected by summing the logical products of the inverted binary image of the current frame and the mask for each pitch and yaw values. The maximum value among the summed logical products corresponds to the current pose of the eye.

This method was initially proposed in (8) and (5). The current implementation moves toward a maximum using a Greedy algorithm. Note that the reflections on the iris area can result in white spots on the binary image that may introduce errors when computing the sum of the logical products.

EXPERIMENTAL RESULTS

This section presents the experimental results obtained using a high-sensitivity camera in low-light condition as well as prototype glasses. Our solution is implemented using OpenCV 3.1 C++ functions wrapped by a Python frontend. OpenCV CUDA modules are called whenever possible to benefit from GPU acceleration.

Object Recognition

In order to detect the focused object, we combine the gaze direction obtained from the tracking with object recognition from cornea reflections. We propose a method based on matching 2D features between the corneal images and a reference image:

- First, we detect features and extract descriptors using Speeded-Up Robust Features (SURF).
- Then, we match the descriptor vectors using Fast Library for Approximate Nearest Neighbors (FLANN).
- Finally, we remove outliers using Random Sample Consensus (RANSAC).

Figure 5 shows the result. The detection is not perfect and suffers from noise due to iris contamination and distortions of the reflection. Further strategies must be applied to isolate the object from incorrect matches. However, our current implementation using this method runs in real-time.

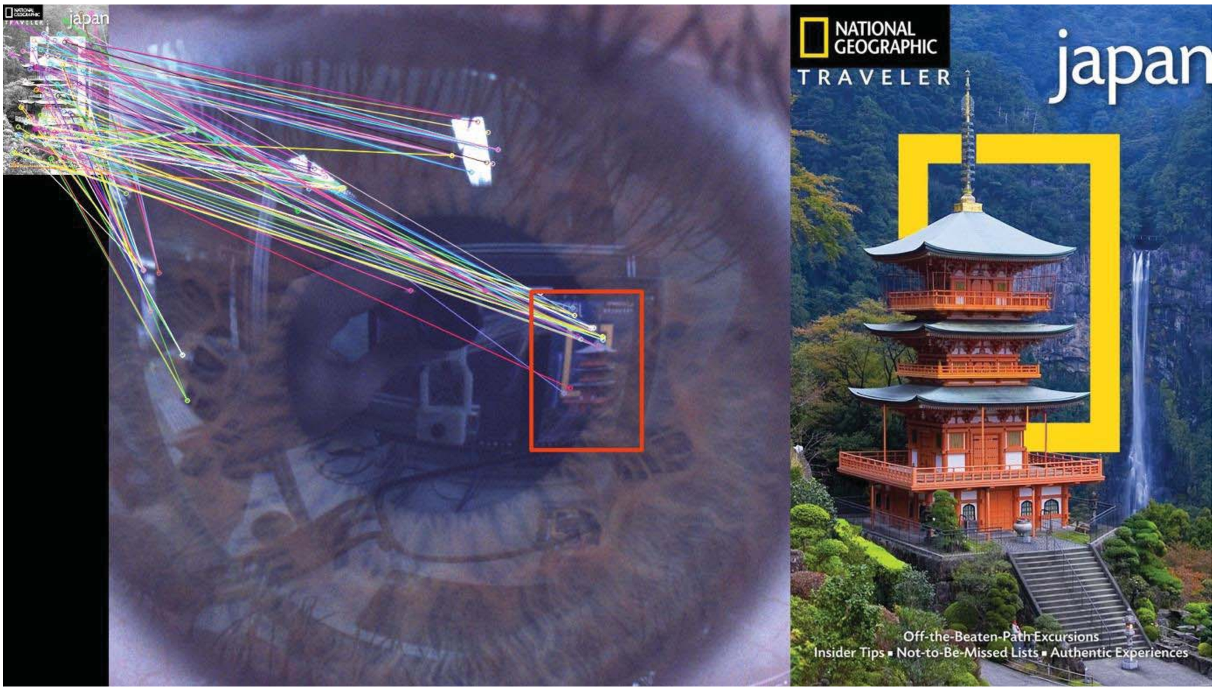

Figure 5 – Object recognition using 2D feature matching

High Dynamic Range

Cornea images are highly sensitive to light conditions and relying exclusively on a single RGB camera at night is challenging. To address this issue, we conducted an experiment using a ViewPLUS Xviii high-sensitivity camera (9) and High Dynamic Range (HDR) processing at night time. The Xviii camera captures eleven 8-bit RGB images at increasing levels of sensitivity over an 18-bit dynamic range. We combine these images using a HDR algorithm, such as Exposure Fusion, to reveal the features reflected on the cornea. Figure 6 shows the result of a user watching a computer screen (displaying the reference object of Figure 5) in the darkness. By using a high-sensitivity camera combined with HDR techniques, we are able to apply our proposed gaze-tracking method in low-light condition.

Figure 6 – HDR processing result (right) in low-light condition

Prototype Glasses

To assess the precision of our tracking method, we prototyped a head-mounted device using a pair of JINS MEME glasses (10) as a base. We mounted on top of them a 3D-printed frame to fix the cameras used for corneal imaging, as shown in Figure 7. We use two e-con Systems See3CAM_80 RGB cameras (11). Eye images from both eyes can be captured at a near-4K resolution of up to 3264 × 2448 pixels at 11 frames per second. The two video streams are passed to a computer through two USB 3.0 cables via an USB Video Class (UVC) 1.1 standard interface.

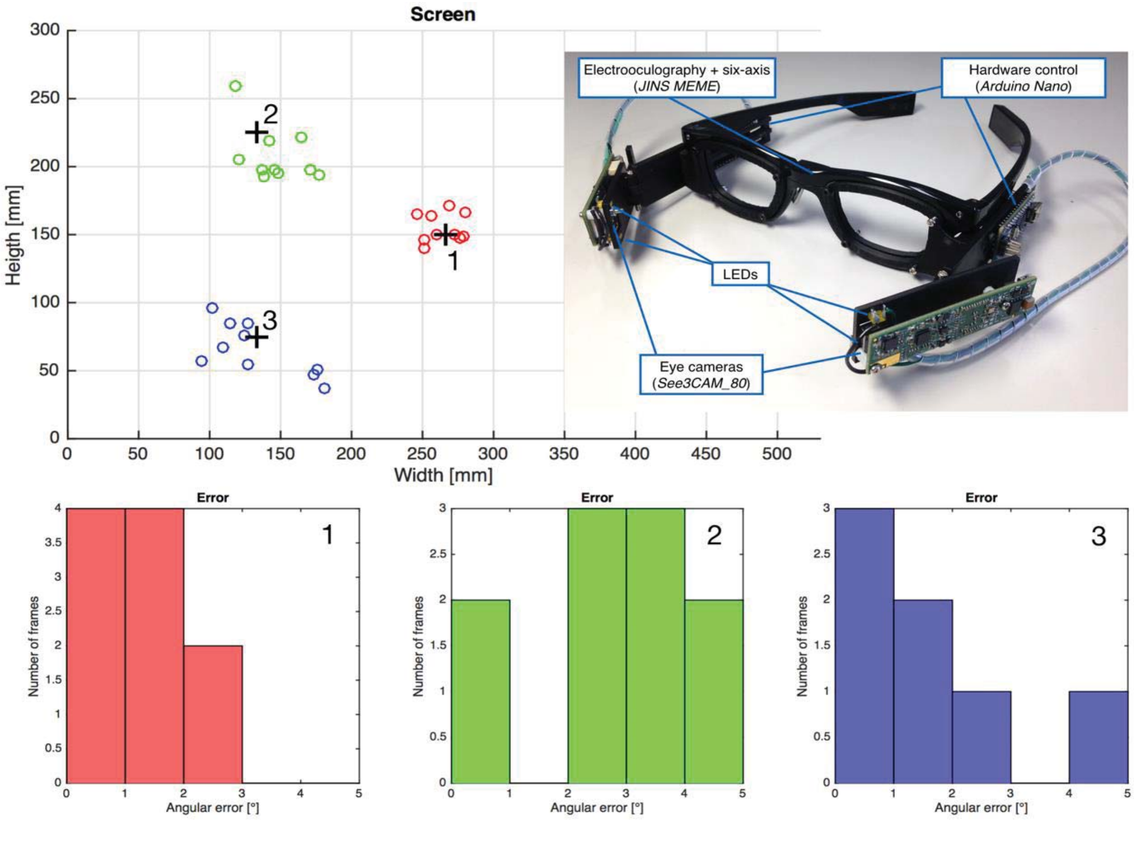

Figure 7 – Experimental results using prototype glasses

During the experiment, a user is asked to sit in front of a computer screen and look at several targets displayed at each corner while staying still, as shown in Figure 7. The camera position relative to the screen is measured and assumed to be constant. The visual axis direction is computed from the measurements and compared with the one given by the implementation of the proposed method.

Results obtained with a screen placed at 500 mm indicate a mean error of 2.5° and a maximum error of 5°. Most commercial infrared-based solutions advertise higher precision, usually below 1°. However, they exclusively focus on gaze direction estimation and require additional sensors to extract scene information. Moreover, they do not work well outdoors because of the sun. Hence, targeting an angular error of less than 3° seems reasonable when using natural images prone to more noise than infrared ones, especially considering that extracting scene information at the same time as gaze direction could make up for the lack of precision, depending on the application.

CONCLUSION

This work explored the feasibility of wearable eye-tracking systems solely relying on a single camera to evaluate the direction to the POR and recognize objects reflected on the cornea surface. However, cornea images are highly sensitive to light condition and many difficulties remain, such as the calibration of the visual axis and the individual variations of the 3D model parameters. Future work will include:

- Extracting the anatomical parameters of the user to improve accuracy.
- Removing the extra step of visual axis calibration.
- Evaluating not only the direction of the gaze, but also the depth to the POR. This could be achieved with stereo reconstruction from left and right cornea images.
- Unwrapping cornea images from the reflected POR to remove distortions before applying object recognition algorithms.

Hence, eye tracking based on corneal imaging requires further investigations to address these challenges in order to become convenient enough for daily life purposes.

REFERENCES

(1) M. Geuss, Why eye tracking could make VR displays like the Oculus Rift consumer-ready, Ars Technica, 2014.

(2) C. Nitschke, A. Nakazawa, and H. Takemura, Corneal Imaging Revisited: An Overview of Corneal Reflection Analysis and Applications, IPSJ Transactions on Computer Vision and Applications, vol. 5, pp. 1–18, 2013.

(3) K. Nishino and S. K. Nayar, Corneal Imaging System: Environment from Eyes, International Journal of Computer Vision, vol. 70, no. 1, pp. 23–40, 2006.

(4) C. Nitschke, A. Nakazawa, and H. Takemura, Display-camera calibration using eye reflections and geometry constraints, Computer Vision and Image Understanding, vol. 115, no. 6, pp. 835–853, 2011.

(5) K. Takemura, T. Yamakawa, J. Takamatsu, and T. Ogasawara, Estimation of a focused object using a corneal surface image for eye-based interaction, Journal of Eye Movement Research, vol. 7, no. 3, pp. 1–9, 2014.

(6) L. El Hafi, K. Takemura, J. Takamatsu, and T. Ogasawara, Model-Based Approach

for Gaze Estimation from Corneal Imaging Using a Single Camera, Proceedings of IEEE/SICE International Symposium on System Integration (SII), 2015, pp. 88–93.

(7) D. Eberly, Computing a Point of Reflection on a Sphere, pp. 1–4, 2008.

(8) K. Takemura, S. Kimura, and S. Suda, Estimating point-of-regard using corneal surface image, Proceedings of Symposium on Eye Tracking Research and Applications (ETRA), 2014, pp. 251–254.

(9) ViewPLUS, Xviii, http://www.viewplus.co.jp/product/camera/xviii.html, 2015.

(10) JIN, JINS MEME, https://jins-meme.com/, 2015.

(11) e-con Systems, See3CAM_80, https://www.e-consystems.com/8MP-USB3-Autofocus-Camera.asp, 2013.

(12) L. El Hafi, P. Uriguen Eljuri, M. Ding, J. Takamatsu, and T. Ogasawara, Wearable Device for Camera-Based Eye Tracking: Model Approach Using Cornea Images, Proceedings of JSME Robotics and Mechatronics Conference (ROBOMECH), 2016.

ACKNOWLEDGEMENTS

This work was conducted under the MEXT scholarship program of the Ministry of Education, Culture, Sports, Science and Technology of Japan.