Estimating Focused Object using Corneal Surface Image for Eye-based Interaction

Kentaro Takemura
Tokai University

Tomohisa Yamakawa
Nara Institute of Science and Technology

Jun Takamatsu
Nara Institute of Science and Technology

Tsukasa Ogasawara
Nara Institute of Science and Technology

In head-mounted camera systems such as Google's Project Glass, researchers are considering the use of eye tracking. Typical methods require a detailed calibration in advance, and long periods of use break the record of calibration between the eye and the scene camera. Additionally, even if the point-of-regard is estimated using a portable eye-tracker, the focused object cannot be estimated. Therefore, we propose a novel method to estimate the object that the user is focusing on, by using an eye camera that captures the reflection on a corneal surface. Eye and environment information can be extracted simultaneously from the corneal surface image. We use inverse ray tracing to rectify the reflection image and a scale-invariant feature transform to estimate the object on which the point-of-regard is located. We believe that our proposed method can be applied to a guide system, and we confirm the feasibility of such usage through experiments that estimate the object focused on and the point-of-regard.

Keywords: Corneal Surface Image, Point-of-regard, Focused object

Introduction

A head-mounted camera system such as Project Glass recently released by Google requires an eye-based intuitive input method. This is also the case with the products of some eye-tracking research, such as the wearable EOG goggles [1] and Aided Eyes [2] that have been proposed as daily-use devices. In general, eye-tracking systems are used to estimate point-of-regard, but it is also necessary, for eye-based interaction, to identify the object that the user is focusing on. In the case of a non-contact eye tracker for the estimation of points-of-regard on a display and on an advertisement, the coordinates of a point-of-regard can be converted directly to the focused object. This is because the relationship between the sensor coordinate system and the object coordinate system is fixed. In contrast, when a head-mounted eye tracker is used, it is difficult to determine the focused object because the coordinate system between the user and the object changes dynamically, according to the user's movements. Additionally, ease of calibration and low weight are necessary features in an eye tracker that is intended for regular use. A typical head-mounted eye tracker consists of a scene and eye cameras; hence, a calibration procedure is required prior to use. Nevertheless, if prior calibration is performed completely, the relationship between two cameras might be broken during a long period of use.

One effective solution for these problems is to obtain eye and environmental information simultaneously with only an eye camera. Nishino and Nayer [3] proposed a method for the extraction of environmental information from corneal surface images, and Nitschke and Nakazawa [4] achieved super-resolution image from the corneal surface image. However, in neigher study was a wearable camera employed to capture corneal surface images; hence, it was difficult to apply eye-based interaction. Therefore, we propose a user-friendly head-mounted eye camera system that can extract scene and eye information from corneal images of the user.

Further, we believe that the proposed system can be used in guidance systems in, e.g., a museum, as an intuitive human interface. For such applications, we confirm that the objects that a user focuses on can be estimated from corneal surface images.

The rest of this paper is structured as follows: Section II describes a wearable camera system that has been developed for daily use. The method for extracting scene images from corneal surface images is described in Section

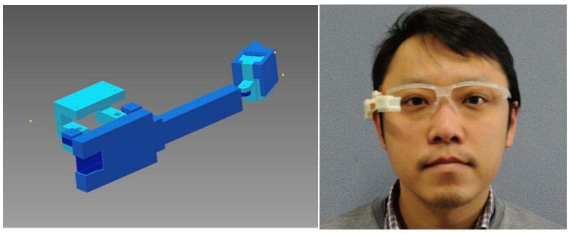

Figure 1. *Portable eye camera for estimating the object being focused using corneal surface image*

Figure 2. *Two prototype systems are developed with a focus on portability.*

III. The methods for the detection of the focused object are described in Section IV and Section V. Section VI, the last section, consists of conclusions and a description of our future work.

Wearable device for capturing corneal surface image

In related works, corneal surface images are captured by a digital single-lens-reflex camera or an industrial camera that can capture high-resolution images [3][4]. The camera is fixed on the floor on a tripod, an arrangement that makes it difficult to get the image sequence and apply it to user interfaces. To avoid this difficulty, we have developed a portable device for the capture of corneal reflections; this is shown in Figure 1. We designed the device using a 3D CAD system, and it was printed on a 3D printer. Our device is different from the conventional eye-tracking systems in that it requires neither a scene camera nor near-infrared LEDs. In addition, the device has a prism installed; this provides enough depth of field for focusing on the eye. We employed a color camera to capture corneal surface images, thus allowing for the use of color information in for image processing. The camera is a small UVC camera, which can therefore be connected to a variety of computers, with the capacity to shoot pictures at 1280×1024 pixels resolution. At present, our algorithm cannot be implemented to operate in real-time—a point we discuss below; hence, the central emphasis in the design of the system configuration was on portability.

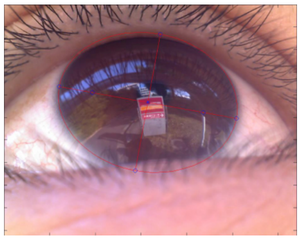

Figure 3. *Iris area is extracted using an ellipse*

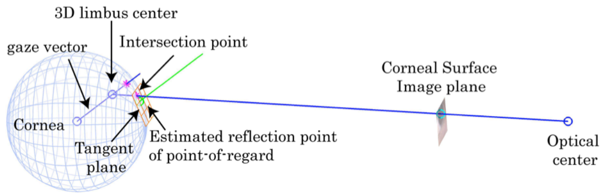

Figure 4. *Relationship between cornea, tangent plane, gaze vector, and point-of-regard*

Figure 2 shows two prototype systems in which Nexus 7 and BeagleBoard-xM are used for the processors. Because both systems lack sufficient storage capacity for the collection of image sequence, we implemented software to upload images to a server with more storage. With this enhancement, corneal surface images can be recorded continuously, at a rate of 2[FPS], until the battery gets discharged.

Extracting scene image from reflection image on corneal surface

Our system has applications in navigation systems that use visual and audio information (e.g., guidance systems in museums). Therefore, it is necessary to identify the object that the user is focusing on. Computer vision researchers have developed various methods for specific object recognition, but these require that the corneal surface images be corrected for distortion before the specific object recognition methods can be applied. The use of 3D eye models in eye tracking has been proposed in recent research [5], and Tsukada et al. [6] proposed a method for the automatic acquisition of the 3D models. With these models, it is possible to define the corneal surface as a part of a sphere and thus generate unwarped images from corneal surface images, Accordingly, we have employed the 3D eye model proposed by Nitschke et al. [7] to estimate the cornea and gaze direction. When the iris is detected, as shown in Figure 3, the relationship between the image plane and the 3D eye model is determined as shown in Figure 4.

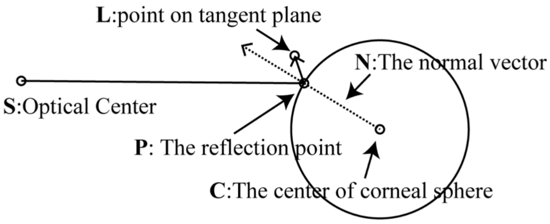

Figure 5. *Geometry model of reflection on corneal surface*

Figure 6. *Result of generating unwarped corneal surface image.Left and right images are input image and generated unwarped corneal surface image respectively.*

However, the distance to the point-of-regard is not known; hence, the reflection point of point-of-regard cannot be computed. Therefore, we have assumed the distance of point-of-regard to be at infinity and instead computed a vector that is parallel to the gaze vector. The unwarped image is generated from the tangent plane at the intersection of the computed parallel vector and corneal sphere, with a specular reflection model and inverse ray tracking employed to generating the image.

As shown in Figure 5, the center **C** of corneal sphere is defined as the point of origin. **S** is the optical center of camera, and **L** is the point of tangent plane. The constraint on specular reflection is calculated from the equation

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

where $a = \mathbf{S}\cdot\mathbf{S}$, $b = \mathbf{S}\cdot\mathbf{L}$, $c = \mathbf{L}\cdot\mathbf{L}$ and $d = |\mathbf{S}\times\mathbf{L}|^2$ are the coefficients of the biquadratic equation.

A normal vector $\mathbf{N} = x\mathbf{S} + y\mathbf{L}$ is computed from a solution to the biquadratic equation, subject to the conditions $x > 0$ and $y > 0$. The unwarped image is generated by inverse ray tracing from the reflection point, which is computed from the equation

$$\mathbf{P} = r_c\mathbf{N} / \|\mathbf{N}\|$$

where r_c is the radius of the corneal sphere.

Figure 6 shows the result of correction the distortion in an image; the object of focus (e.g., human face) is extracted continuously using the head-mounted device.

Figure 7. *Result of color correction. Left and right images are original unwarped corneal image and color-corrected unqorped image.*

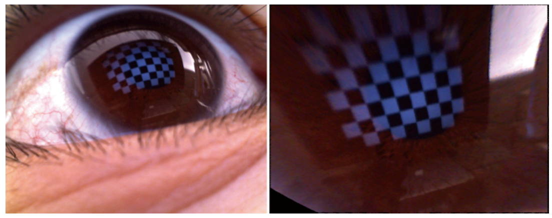

Figure 8. *The corneal surface image which chess pattern is reflected on, and the unwarped corneal surface image is generated for confirming linearity and parallelism.*

The generated image is largely determined by iris color, as shown in Figure 6. As we discuss below, algorithms for estimating focused object depend on intensity alone; so color correction is not indispensable for eye-based interaction. On the other hand, if the proposed system is regarded as a measurement system, then color correction is required to improve visibility. Various algorithms have been developed for achieving color constancy; a well-known algorithm is the Gray World Assumption [8]. The result of applying this algorithm to an unwarped corneal surface image is shown in Figure 7, and the influence of iris color could be largely eliminated.

We have generated a special unwarped image to confirm of linearity and parallelism. The left image in Figure 8 was captured when a user looked at a chess pattern on a display, and right image was generated as the unwarped image. That the distortion has been corrected can be confirmed from the unwarped image.

Specific object recognition using unwarped corneal surface images

Ocular motion and point-of-regard are important in conventional eye-tracking systems. It is, however, difficult to use the information directly in user interfaces and human-computer interaction in the case of a head-mounted eye tracker. To apply the eye-tracking system to navigation and guidance systems, it is necessary to identify the object the user is focusing on. To do so, we employed a

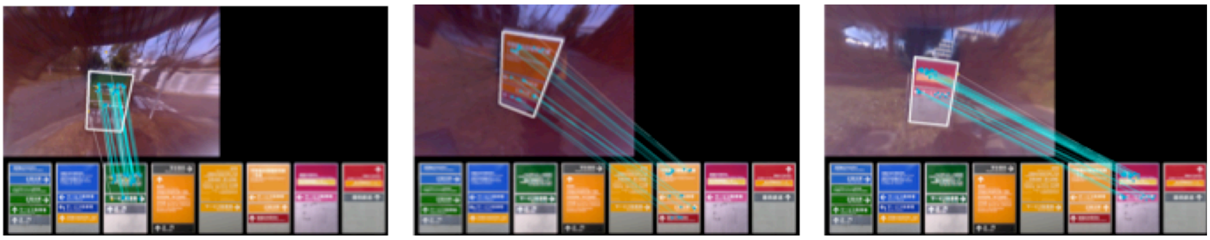

Figure 9. *Results of recognizing focused objects using SIFT feature.*

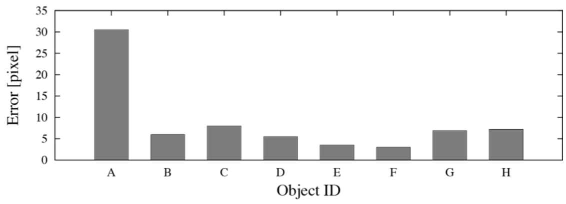

Figure 10. *Average of correspondence errors between registered templates and unwarped corneal surface images.*

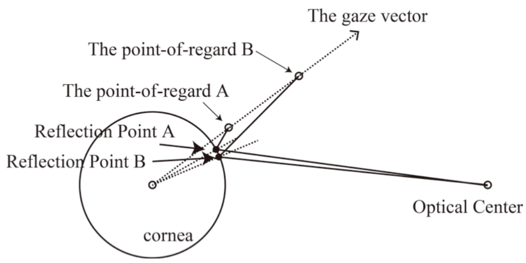

Figure 11. *Reflection point depends on the distance to point-of-regard. Point-of-regard A and B were reflected on different positions respectively.*

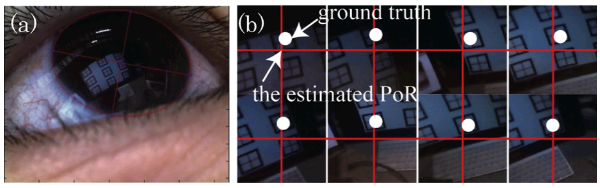

Figure 12. *A snapshot of corneal surface image(a), and the results of estimating point-of-regard using reflection point of vector parallel to a gaze vector(b).*

simple technique for specific object recognition and then conducted an experiment to evaluate the results of specific object recognition with unwarped corneal surface images. Eight outdoor direction boards located around the university were used in the evaluation; corneal images were obtained from a participant looking at these boards. Figure 9 shows the results of specific object recognition using the SIFT (scale invariant feature transform) feature [9]. The white rectangle on each image is the estimated area of the direction board using RANSAC [10].

The average, $\bar{E}$, correspondence errors between the registered template and the unwarped corneal surface image is calculated using the equation

$$\bar{E} = \frac{1}{n}\sum_{i=1}^{n}\left\|\mathbf{HP}_i - \mathbf{P}_i'\right\|$$

where $\mathbf{P}_i$ is a feature of the unwarped corneal surface image, and $\mathbf{P}_i'$ is a feature of the registered template. The homography matrix, $\mathbf{H}$, is calculated from four points selected manually on the registered image and the unwarped corneal surface image. Figure 10 shows the correspondence errors from the eight direction boards. On the whole, matching was achieved with high accuracy, but with average error in the case of object A; the main cause of this error was the existence of similar characters in object A. Nevertheless, we did confirm the feasibility of specific object recognition with an unwarped corneal image.

Extracting focused object using point-of-regard

A focused object can be recognized by SIFT features, as previously mentioned. If, however, there are two objects on a corneal surface image, the focused object cannot be detected; in such a case, it is important to set the tangent plane appropriately using the point-of-regard. However, as shown in Figure 11, the point-of-regard cannot be estimated correctly unless the distance to focused object is known. What we have done, as stated above, is assume that the point-of-regard is infinity and then estimate its actual location on the basis of a vector that is parallel to the gaze vector. Beyond that, we compared the estimated point-of-regard and with the ground truth; Figure 12 summarizes this. Points-of-regard were estimated when a user looked at nine crosses on the display and ground truths were identified pointed manually. A point-of-regard is always estimated to be slightly below the ground truth in the image plane. The estimate error is

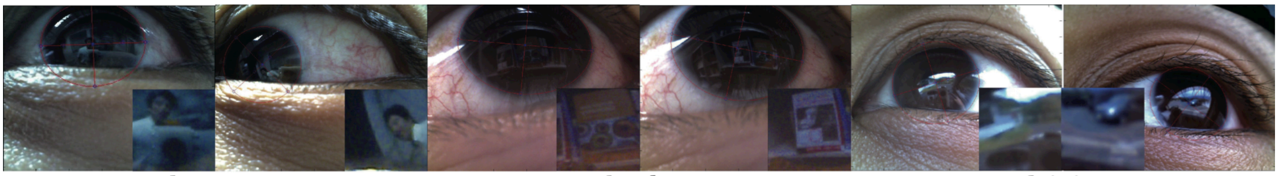

Figure 13. *The results of extracting focused object based on point-of-regard under various situations*

about 8[deg] when a user focuses on an object at a distance of 2.5 [m]. In other words, the error of the point-of-regard is within 350 [mm], when the user stands 2.5[m] from the focused object. As our goal is to achieve eye-based interaction such as a guidance system that has sufficient accuracy to extract the focused object. Figure 13 shows the result of extracting focused object, on the basis of points-of-regard, under various situations. In the first case, the image is of that of the user's cornea on meeting a friend, and the second case is that of the user's cornea when looking at two books; the cornel image is extracted correctly in both cases. We are nevertheless aware of two problems with the proposed system. One of these is its dependence on illumination condition: a reflected image might be reasonably clear outdoors, but it is difficult to get a clear image indoors unless the corneal image is captured under good illumination. The other problem is the distance to point-of-regard: in driving a vehicle, a driver looks into the far distance; the resulting image will be lacking in resolution, and this makes it difficult to recognize the focused object.

Conclusions

We developed a wearable device for capturing corneal surface images for daily-use devices, and we showed specific object recognition could be achieved with unwarped corneal surface images. We also confirmed the feasibility of our approach through preliminary experiments.

In the future, we plan to implement a portable navigation system based on the estimated focused object. This will require the solution of additional problems, such as how to ensure low processing time and high image quality.

Acknowledgements

This research was supported by the Strategic Information and Communications R&D Promotion Programme of the Ministry of Internal Affairs and Communications, Japan.

References

1. Bulling, A., Roggen, D. & Tröster, G. (2009). *Wearable EOG Goggles: Seamless Sensing and Context-awareness in Everyday Environments*. Journal of ambient Intelligence and Smart Environments. vol. 1, no.2, pp.157-171.
2. Ishiguro,Y., Mujibiya, A., Miyaki, T. & Rekimoto, J. (2010). *Aided eyes: eye activity sensing for daily life.* Proc. of the 1st Augmented Human International Conference, pp. 1–7.
3. Nishino, K. & Nayar, S. K. (2004). *The world in an eye.* Proc. of the 2004 IEEE Computer Society Conference Computer Vision and Pattern Recognition, vol.1, pp.444-451.
4. Nitschke, C. & Nakazawa, A. (2012). *Super-Resolution from Corneal Images.* Proc. of the British Machine Vision Conference, pp.22.1–22.12.
5. Nagamatsu,T., Kamahara J., & Tanaka, N. (2009). *Calibration-free Gaze Tracking Using Binocular 3D Eye Model.* CHI EA '09 Proceedings of the 27th international conference extended abstracts on Human factors in computing systems, pp. 3613-3618.
6. Tsukada, A. & Kanade, T. (2012). *Automatic acquisition of a 3D eye model for a wearable first-person vision device*. Proceedings of the Symposium on Eye Tracking Research and Applications.
7. Nitschke, C., Nakazawa, A. & Takemura, H. (2011). *Display-camera calibration using eye reflections and geometry constraints*. Computer Vision and Image Understanding, vol. 115, no.6, pp. 835-853.
8. Buchsbaum, G. (1996). *A spatial processor model for object colour perception*. Journal of The Franklin Institut, pp.83-88.
9. Lowe, D. G. (2004). *Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision,* 60, 2, pp. 91-110.
10. Fischler, M.A. & Bolles, R.C. (1981). *Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography.* Comm. Of the ACM 24(6), pp.381-395.