Calibration-free Gaze Tracking Using a Binocular 3D Eye Model

Takashi Nagamatsu
Kobe University
5-1-1, Fukae-minami, Higashi-nada, Kobe 658-0022, JAPAN
nagamatu@kobe-u.ac.jp

Junzo Kamahara
Kobe University
5-1-1, Fukae-minami, Higashi-nada, Kobe 658-0022, JAPAN
kamahara@maritime.kobe-u.ac.jp

Naoki Tanaka
Kobe University
5-1-1, Fukae-minami, Higashi-nada, Kobe 658-0022, JAPAN
ntanaka@maritime.kobe-u.ac.jp

Abstract
This paper presents a calibration-free method for estimating the point of gaze (POG) on a display by using two pairs of stereo cameras. By using one pair of cameras and two light sources, the optical axis of the eye and the position of the center of the cornea can be estimated. This estimation is carried out by using a spherical model of the cornea. One pair of cameras is used for the estimation of the optical axis of the left eye, and the other pair is used for the estimation of the optical axis of the right eye. The point of intersection of optical axis with the display is termed the point of the optical axis (POA). The POG is approximately estimated as the midpoint of the line joining POAs of both the eyes with the display. We have developed a prototype system based on this method and demonstrated that the midpoint of POAs was closer to the fiducial point that the user gazed at than each POA.

CHI 2009, April 4 – 9, 2009, Boston, MA, USA
ACM 978-1-60558-247-4/09/04.

Keywords
Gaze tracking, calibration-free, eye movement, eye model

ACM Classification Keywords
H.5.2 [Information Interfaces and Presentation]: User Interfaces–Ergonomics; I.4.9 [Image Processing and Computer Vision]: Applications

Introduction
Gaze tracking technology is used as a human-machine interface [1, 5, 6]. However, most accurate gaze tracking systems require a personal calibration process in which the user is instructed to gaze at 2–9 points

before the system can be used. Generally, the data obtained by the calibration process can be reliably used only in the trial that is conducted after it. In other words, the system must be recalibrated if one of its parameters changes. Because of the calibration process, the application of gaze tracking systems is restricted. It cannot be used to develop systems meant for the general public.

Several studies have attempted to reduce the calibration effort using a model-based approach. Figure 1 shows a typical eye model that is used in a model-based approach[2]. An eye has two axes: one is the optical axis that is the geometric center line of the eye, and the other is the visual axis that is the line of sight connecting the fovea and the point of gaze (POG). In the model, it is approximated that the two axes intersect at the center of the corneal curvature. The horizontal angle between the optical and visual axes ranges from 3.5° to 7.5° (average of 5.5°), and the lateral angle between the optical and visual axes ranges from 0.25° to 3.0° (average of 1.0°) [9].

Shih and Liu [10] and Guestrin and Eizenman [2, 3] presented methods for reconstructing the optical axis of the eyeball using stereo cameras without actually knowing the characteristics of an individual's eye. These methods also estimate the visual axis through a one-point calibration process that requires a user to gaze at only a single point. We improved these methods to precisely calculate the rotation of the eyeball on the basis of Listing's law [7, 8]. If calibration is not conducted, the optical axis is assumed to be an approximation of the visual axis; hence, the accuracy of the approximation is about 5°. Personal calibration has to be conducted to accurately estimate the POG.

In this paper, we present a novel calibration-free gaze tracking method using a display. While previous studies on one-point calibration measured the optical axis of only one eye, we measure the optical axes of both eyes. Our method estimates the offset between the optical and visual axes without requiring the user to intentionally gaze at the fiducial point.

The following section explains the method to estimate the optical axis of a single eye. Next, calibration-free gaze tracking using a binocular 3D eye model is described. In the final section, a prototype system developed on the basis of these methods is described.

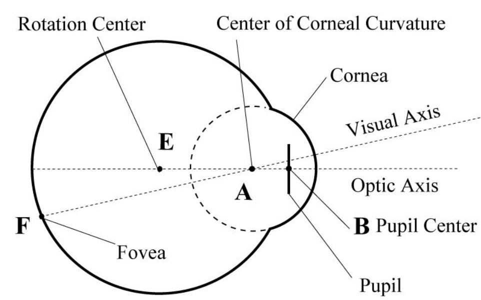

Figure 1. Eye model.

Estimation of optical axis of single eye

The optical axis can be estimated by the method described in [3, 8, 10]. This method requires a minimum of two cameras and two point light sources. The cameras are modeled as pinhole type cameras. The positions of the light sources are measured, and the intrinsic and extrinsic camera parameters are determined beforehand by camera calibration. Points on the camera image plane, such as the first Purkinje images (reflection of point light sources from the outer surface of the cornea) and the center of the pupil, are detected by image processing and converted into a 3D position by setting the camera parameters beforehand.

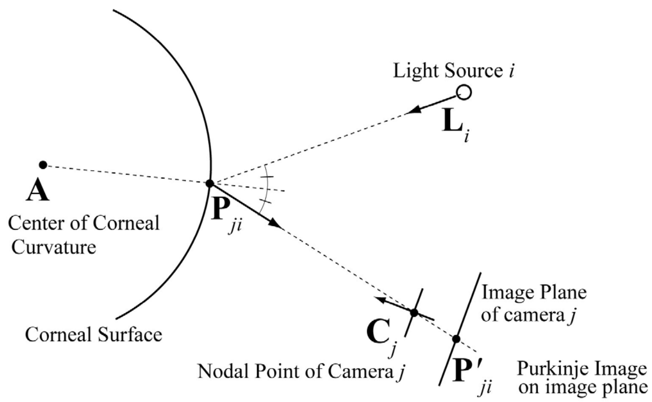

Figure 2. Cross section of the eyeball, showing the center of corneal curvature, light source, and nodal point of the camera.

Estimation of center of corneal curvature

We estimate the position of the center of the corneal curvature, $\mathbf{A}$ (bold face indicates a position vector). Figure 2 shows a cross section of the eyeball where $\mathbf{A}$ is the center of the corneal curvature, $\mathbf{L}_i$ indicates the position of the light source i, and $\mathbf{C}_j$ denotes the nodal point of camera j. The value of $\mathbf{L}_i$ and $\mathbf{C}_j$ are known. A ray from $\mathbf{L}_i$ is reflected at a point $\mathbf{P}_{ji}$ on the corneal surface such that it passes through $\mathbf{C}_j$ and intersects the camera image plane at a point $\mathbf{P}'_{ji}$. $\mathbf{C}_j$, $\mathbf{P}_{ji}$, $\mathbf{L}_i$, $\mathbf{P}'_{ji}$, and $\mathbf{A}$ are coplanar, and the normal vector of the plane is $(\mathbf{P}'_{ji} - \mathbf{C}_j) \times (\mathbf{L}_i - \mathbf{C}_j)$. The plane is expressed as:

$$\{(\mathbf{P}'_{ji} - \mathbf{C}_j) \times (\mathbf{L}_i - \mathbf{C}_j)\} \cdot (\mathbf{x} - \mathbf{C}_j) = 0$$

where $\mathbf{x}$ $(= (x, y, z))$ is a point on the plane. We obtain four planes when we use two cameras and two light sources ($i = 0, 1$ and $j = 0, 1$). All planes contain $\mathbf{A}$, and it can be calculated if at least three planes are given.

Estimation of optical axis

Figure 3 shows a plane that contains $\mathbf{A}$, $\mathbf{B}$ (center of the pupil), and $\mathbf{C}_j$. A ray from $\mathbf{B}$ refracts at $\mathbf{B}''_j$, passes through $\mathbf{C}_j$, and reaches $\mathbf{B}'_j$ (pupil center on the image sensor). Therefore, the plane that includes $\mathbf{A}$, $\mathbf{B}''_j$, and $\mathbf{B}'_j$ also includes the optical axis (line connecting $\mathbf{A}$ and $\mathbf{B}$). Because we use two cameras, we can calculate the optical axis as the intersection of the two planes ($j = 0$, 1). The normal vector of the plane is $(\mathbf{C}_j - \mathbf{B}'_j) \times (\mathbf{A} - \mathbf{C}_j)$.

Therefore, the unit vector of the optical axis, $\mathbf{d}$, can be written as:

$$\mathbf{d} = \frac{((\mathbf{C}_0 - \mathbf{B}'_0) \times (\mathbf{A} - \mathbf{C}_0)) \times ((\mathbf{C}_1 - \mathbf{B}'_1) \times (\mathbf{A} - \mathbf{C}_1))}{|((\mathbf{C}_0 - \mathbf{B}'_0) \times (\mathbf{A} - \mathbf{C}_0)) \times ((\mathbf{C}_1 - \mathbf{B}'_1) \times (\mathbf{A} - \mathbf{C}_1))|} . \qquad (\quad)$$

The optical axis is expressed as $\mathbf{x} = \mathbf{A} + t\mathbf{d}$ (t is a parameter).

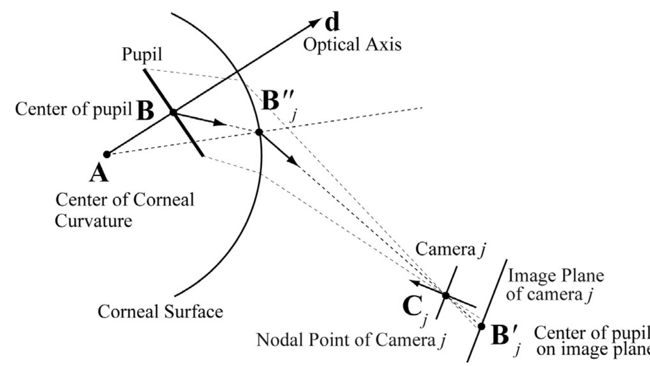

Figure 3. Cross section of the eye ball, showing the center of corneal curvature, pupil center, and nodal point of the camera.

Estimation of intersection of optical axis with display
The display can be expressed by using the following four points: $\mathbf{D}_{\mathrm{TL}}$, $\mathbf{D}_{\mathrm{TR}}$, $\mathbf{D}_{\mathrm{BL}}$, and $\mathbf{D}_{\mathrm{BR}}$. The normal vector of the display is $(\mathbf{D}_{\mathrm{TL}} - \mathbf{D}_{\mathrm{BR}}) \times (\mathbf{D}_{\mathrm{TR}} - \mathbf{D}_{\mathrm{BL}})$; therefore, the display can be expressed as
$$\{(\mathbf{D}_{\mathrm{TL}} - \mathbf{D}_{\mathrm{BR}}) \times (\mathbf{D}_{\mathrm{TR}} - \mathbf{D}_{\mathrm{BL}})\} \cdot (\mathbf{x} - \mathbf{D}_{\mathrm{TL}}) = 0 .$$

The point of intersection of optical axis with the display is termed the point of the optical axis (POA). After estimating the POA in world coordinates, we can convert it into the graphical coordinates of the display.

Calibration-free gaze tracking method

Binocular eye model
Our previous system [8] estimated the POG by one-point calibration. Before a user uses the system, he or she must gaze at one point whose position is known. Therefore, the system can estimate the offset between the optical and visual axes.

The new calibration method can determine the offset between the optical and visual axes by using a binocular optics. As described in the introduction, the average horizontal angle between the optical and visual axes is 5.5°, and the average lateral angle between them is 1.0°. It is known that the visual axis inclines toward the nose, away from the optical axis. Furthermore, when a user gazes at an object on the display, the visual axes intersect at a point on the display. As shown in Figure 4, the midpoint of POAs can be a good approximation of the POG, since the horizontal offsets of the POAs from the POG cancel each other and the average lateral offset is 1.0°.

Although the accuracy of estimation of the POA cannot be measured, it is assumed to be approximately 1.0°, because the accuracy of estimation of the POG after one-point calibration was about 1.0° in our previous study [8]. Therefore, the method is expected to have an accuracy of approximately 2.0°, which is better than the accuracy of the method in which the visual axis is approximated by the optical axis.

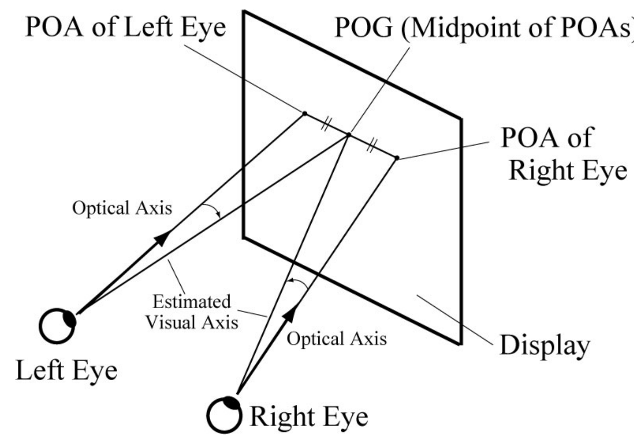

Figure 4. Binocular 3D eye model.

Operational stability
When the system is unable to detect an eye due to image processing problems, the POG is calculated as follows. Once the POG is estimated from the two POAs, the offset between the optical and visual axes is calculated using the one-point calibration method [8] by considering the fiducial point as the midpoint of the POAs. Then, the calculated offset of the eye that the system could detect is used for the estimation of the POG. Thus, the stable operation of the system is achieved.

Implementation and demonstration

Implementation

A prototype system was implemented, as shown in Figure 5. This system consists of four synchronized monochrome IEEE-1394 digital cameras with a 1/3" CMOS image sensor (Firefly MV, Point Grey Research Inc.) that has a 50-mm lens and an IR filter, three infrared light sources attached to a 19" LCD, and a Windows-based PC (Windows XP, Intel Core 2 Quad). The software was developed using OpenCV [4]. A pair of cameras was used for the estimation of the optical axis of the left eye, and the other pair was used for the right eye. Three LEDs were used for precise estimation of the optical axes, because they compensate for the asphericity of the cornea, as described in [8]. Because of the narrow range of view angle and focus of the cameras, the area where a user can move is limited.

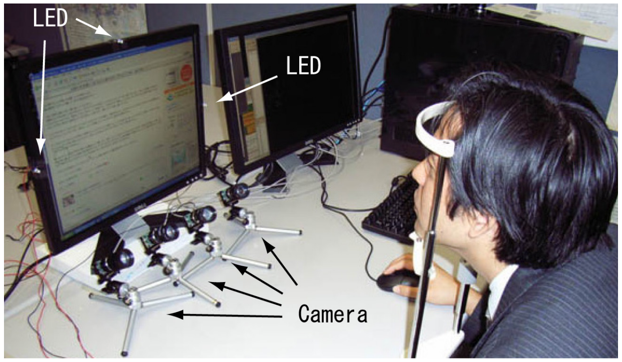

Figure 5. Prototype system.

Demonstration

We demonstrated the prototype system in a laboratory with an adult subject without glasses or contact lenses. The head was supported by a chin rest not to be out of focus of the cameras. The eyes were approximately 600 mm from the display. The subject was asked to fixate on 25 points on the display. Data was recorded when both the optical axes were detected; recording was continued until data of 10 frames was obtained for each point.

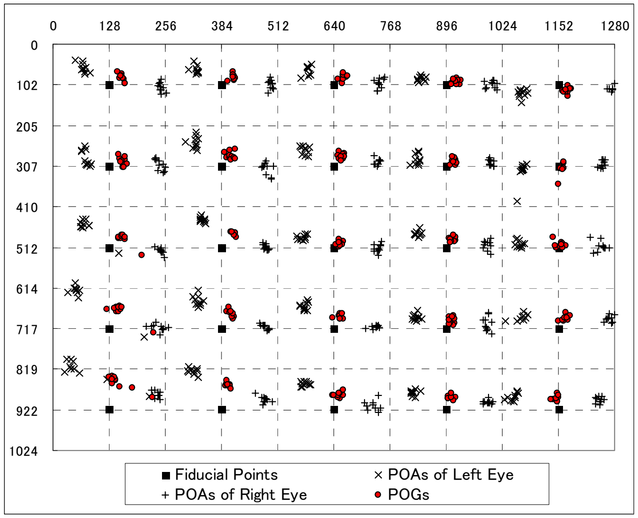

Figure 6. Measured results.

The measured POGs are plotted in the display coordinate system, as shown in Figure 6. The square-shaped points represent fiducial points that were intentionally gazed at by the subject. The x-shaped and +-shaped points indicate the POA of the left and right eyes, respectively. The circular points indicate the estimated POG. The average of the visual angles was 0.91°; the visual angle is defined as the angle between the line connecting a POG with the midpoint of the centers of the cornea and the line connecting the fiducial point with the same midpoint.

Discussion

We found that the midpoint of the line joining the POAs were closer to the fiducial point that the user gazed at than to each POA. However, since the midpoint of the line joining the POAs were calculated, the accuracy of estimating the POG largely depends on the user. Therefore, the experiment must be conduced with more subjects. Furthermore, we intend to develop a method that estimates the POG by actual measurement rather than from the average of POAs. As for the area where a user can move, it can be improved by using a high resolution camera with a wide-angle lens and/or a pan-tilt unit with head tracking.

Conclusion

A calibration-free gaze tracking method has been proposed. It uses two pairs of stereo cameras. One pair of cameras is used for the estimation of the optical axis of the left eye, and the other pair is used for the estimation of the optical axis of the right eye. The POG is estimated as the midpoint of the line joining the POAs of both the eyes. We developed a prototype system and demonstrated that the midpoint of the line joining the POAs were closer to the fiducial point that the user gazed at than to each POA.

Acknowledgements

This research was partially supported by the Japan Science and Technology Agency, Research for Promoting Technological Seeds, 2008.

References

[1] Duchowski, A.T. *Eye Tracking Methodology: Theory and Practice, 2nd ed* Springer-Verlag, London (2007).

[2] Guestrin, E.D. and Eizenman, M. General Theory of Remote Gaze Estimation Using the Pupil Center and Corneal Reflections *IEEE Transactions on Biomedical Engineering*, *53* (6) (2006), 1124-1133.

[3] Guestrin, E.D. and Eizenman, M. Remote Point-of-Gaze Estimation with Free Head Movements Requiring a Single-Point Calibration. in *Proceedings of the 29th Annual International Conference of the IEEE EMBS* (2007), 4556-4560.

[4] Intel Open Source Computer Vision Library. http://www.intel.com/technology/computing/opencv/index.htm

[5] Jacob, R.J.K. The use of eye movements in human-computer interaction techniques: what you look at is what you get *ACM Transactions on Information Systems*, *9* (2) (1991), 152-169.

[6] Nagamatsu, T., Kaieda, Y., Kamahara, J. and Shimada, H. Development of a Skill Acquisition Support System Using Expert's Eye Movement. in *Proceedings of HCI International 2007*, *9* (2007), 430-439.

[7] Nagamatsu, T., Kamahara, J., Iko, T. and Tanaka, N. One-point Calibration Gaze Tracking Based on Eyeball Kinematics Using Stereo Cameras. in *Proceedings of Eye Tracking Research & Applications Symposium 2008* (2008), 95-98.

[8] Nagamatsu, T., Kamahara, J. and Tanaka, N. 3D Gaze Tracking with Easy Calibration Using stereo Cameras for Robot and Human Communication. in *Proceedings of IEEE RO-MAN 2008* (2008), 59-64.

[9] Osaka, R. *Experimental psychology of eye movements (in Japanese)* The University of Nagoya Press., Nagoya,Japan (1993).

[10] Shih, S.-W. and Liu, J. A novel approach to 3-D gaze tracking using stereo cameras *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, *34* (1) (2004), 234-245.