Inside-Out Camera for Acquiring 3D Gaze Points

Hironobu Fujiyoshi[1] Yuto Goto[1] Makoto Kimura[2]
Department of Computer Science College of Engineering Chubu University[1]
Samsung Yokohama Research Center[2]
hf@cs.chubu.ac.jp[1], m.kimura@samsung.com[2]

Abstract

First Person Vision, which attempts to understand a person's behavioral intention, requires information on the person's state and on what the person is looking at. We propose an inside-out camera that simultaneously obtains an image of one of the person's eyeballs and an image of that person's visual field, and propose a method for estimating the person's gaze point based on the configuration of the camera. The relationship between the gaze vector obtained from the eyeball video and the gaze point obtained from the scene video is expressed by a conversion equation. In an evaluation experiment, we took scene video for both eyes when gazing at a certain landmark as evaluation data, and calculated the error between the position of the landmark in the scene video as the true value and the gaze point estimated from the gaze vector.

1. Introduction

First Person Vision attempts to understand a person's behavioral intentions by attaching cameras to a person's head, one to observe one of the person's eyeballs and another to simultaneously obtain an image of the corresponding visual field [1].

The First Person Vision approach uses a camera to detect the person's gaze with the aim of understanding the behavioral intentions of that person.

With the above in mind, we propose a camera system that we call the "inside-out camera" that uses half-mirrors to capture the eyeball head-on and the visual field from a position nearly equivalent to the eyeball. We also propose a technique for calculating the person's gaze point by exploiting the structural features of the inside-out camera. This camera system consists of four cameras in total, two eye cameras for capturing the left and right eyeballs and two scene cameras for capturing the corresponding visual fields. To calculate the gaze point in this system, we estimate beforehand the parameters of an equation for converting the gaze vector estimated from the eye-camera video to a gaze point that exists in the scene-camera video. These parameters enable the gaze point to be calculated from a motion vector.

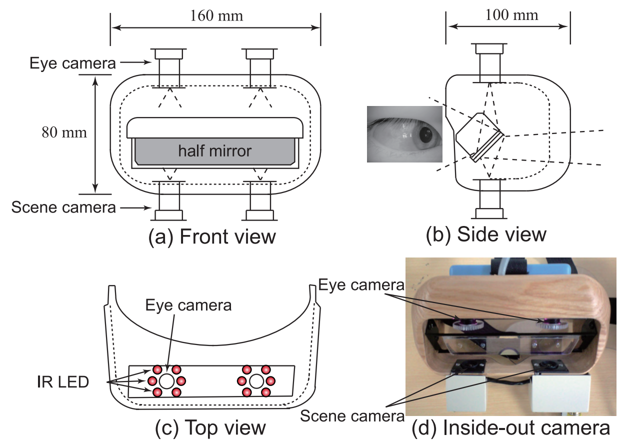

Figure 1. Inside-out camera system

2. Inside-Out Camera

Our most recent prototype equipment has the shape of goggles as shown in Fig. 1. It consists of two eye cameras installed at the top of the unit for capturing images of the subject's eyeballs and two scene cameras installed at the bottom of the unit for capturing the subject's visual field. The equipment measures W160xH80xD100 mm and weighs about 200 g. It is made of wood to hinder the conduction of heat emitted by the camera system to the subject and to enable measurements to be performed over a relatively long time. As shown in Fig. 2, the inside-out camera achieves an optical configuration in which transparent cameras seem to exist. The following describes the eye and scene cameras in more detail.

2.1. Eye camera

The eye camera system consists of an infrared mirror, two infrared cameras for capturing the left and right eyeballs, and six infrared LEDs arranged around each camera. Each infrared camera captures a near-infrared image of the person's eyeball from in front of that eye via an infrared mirror at a resolution of 640x480 pixels. The LEDs arranged around each camera emit near-infrared light in the

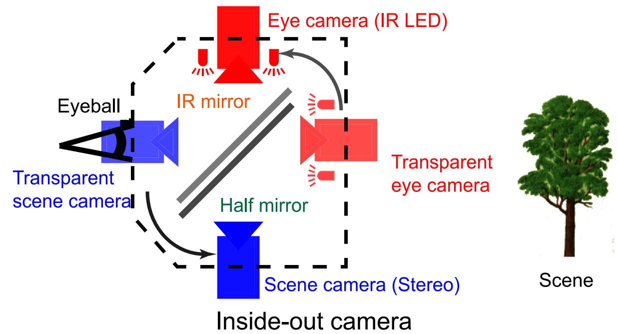

Figure 2. Comparing conventional camera and inside-out camera

wavelength range of 750 - 900 nm. Since infrared light is invisible, it provides no visual stimuli enabling images of the eyeballs to be captured unhindered.

2.2. Scene camera

The scene camera system consists of a 50% half mirror and two compact CCD cameras for capturing the left and right visual fields. The viewing angle of each CCD camera is about 80 degrees, and the focal length is about 4 mm. The 50% half mirror reflects 50% of incident light and allows the rest to pass. The use of a half mirror in this way makes it possible to capture images by a transparent camera from a position that is optically nearly the same as the person's viewing point. Furthermore, as this is a stereo camera system, it is relatively easy to calibrate it using the Tsai model or Zhang model and to estimate the 3D position of the gaze point in the visual field.

2.3. Relationship between eye camera and scene camera

The eye cameras and scene cameras are placed opposite each other with half mirrors in between. The image planes configured by each type of camera are therefore parallel to each other. Now, for an object observed by an eye camera that moves in a similar manner to an object observed by the scene camera, it is clear that a correlation exists between the distance moved by the object observed in the eye-camera video and the distance moved by the object observed in the scene-camera image. The relationship between these two types of cameras is therefore easy to work with.

3. Estimation of 3D Gaze Point using the Inside-out Camera

In general, a gaze vector is needed to estimate the gaze point, and various techniques have been proposed to estimate this vector [5, 6, 2, 7]. With the proposed inside-out camera, this correlation is easy to work with as described in section 2.3. Thus, since a correlation exists between the gaze point and gaze vector, estimating the gaze vector enables the gaze point in the scene image to be estimated. The process of gaze-point estimation has the following flow:

1. Estimate the gaze vector
2. Calculate gaze point
3. Calculate 3D gaze point

3.1. Estimation of gaze vector

Our proposed technique also estimates the gaze vector from the cornea curvature center and pupil center. The inside-out camera that we propose features six light sources on the periphery of the camera, and we can assume that the center of these light sources corresponds to the optical axis of the camera. The center of the Purkinje-image group can therefore be taken to be the cornea curvature center. The cornea curvature center $\boldsymbol{C} = [C_u, C_v]^T$ on the video image can be estimated from the group of Purkinje images determined in this way.

Next, the pupil center is estimated. We use the technique proposed by Sakashita et al. to calculate pupil center $\boldsymbol{P}$ [4].

The gaze vector can now be calculated from the cornea curvature center $\boldsymbol{C}$ and pupil center $\boldsymbol{P}$ estimated as described above. Gaze vector $\boldsymbol{V} = [V_u, V_v]^T$, whose base is taken to be cornea curvature center $\boldsymbol{C}$, can be calculated by the following equation:

$$\boldsymbol{V} = \boldsymbol{P} - \boldsymbol{C} \tag{1}$$

3.2. Calculate gaze point using conversion equation

The gaze point on the scene video can be calculated using the gaze vector estimated in section 3.1 and a conversion equation. This technique is divided into offline processing for determining parameters of this conversion equation and online processing for estimating the gaze point using the conversion equation. This process flow is shown in Fig. 3 and described below.

3.2.1 Offline processing

Offline processing estimates the parameters used for expressing the relationship between the gaze vector and gaze point using a conversion equation. The distribution of u, v components for gaze point $\boldsymbol{L} = [L_u, L_v]^T$ and gaze vector $\boldsymbol{V} = [V_u, V_v]^T$. is shown in Fig. 3(a). It can be seen here that the gaze point and gaze vector have a proportional relationship for each of the u, v components enabling a linear conversion to be performed. The equations for this linear conversion are given below:

$$L_u = a_u V_u + b_u \tag{2}$$
$$L_v = a_v V_v + b_v \tag{3}$$

Here, $\boldsymbol{a} = [a_u, a_v]^T$ is the slope and $\boldsymbol{b} = [b_u, b_v]^T$ is the intercept of these u, v linear equations. Thus, by calculating beforehand the slope and intercept of these equations from at least two calibration points as shown in Fig. 3(b), gaze point $\boldsymbol{L}$ can be calculated on input of gaze vector $\boldsymbol{V}$.

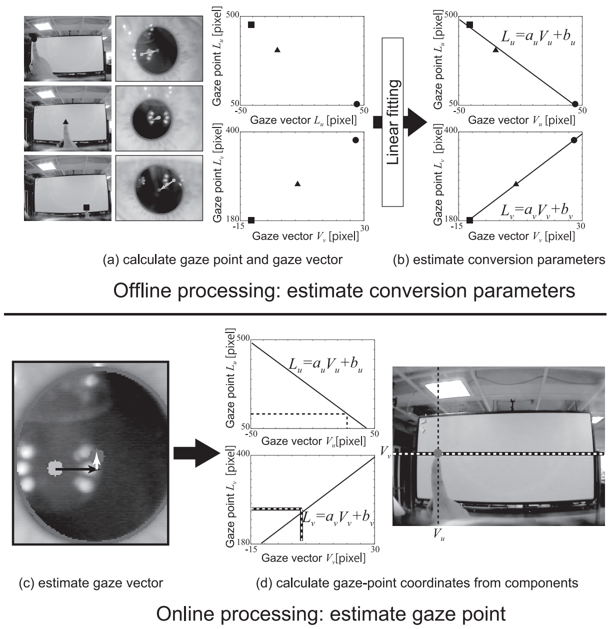

Figure 3. Relation between gaze point and gaze vector

3.2.2 Online Processing

Online processing estimates the gaze point from a gaze vector using conversion parameters $\boldsymbol{a}$, $\boldsymbol{b}$ calculated in offline processing. First, the gaze vector is calculated by the technique described in section 3.1. Next, the calculated gaze vector $\boldsymbol{V}$ is divided into its u, v components and the gaze point is calculated from Eqs.(2) and (3). An example of calculating the gaze point is shown in Fig. 3(d).

3.3. Calculation of 3D gaze point

Finally, we calculate the 3D gaze point using the gaze points of both eyes. Since these gaze points are points on two scene videos, the 3D gaze point can be calculated as a problem in stereo vision. Given that human viewing angle during gazing is about two degrees [3], we calculate a 3D position from the regions of both gaze points.

4. Estimation of 3D gaze point

Figure 4 shows three examples of estimating 3D gaze points from calculated gaze points. In Fig. 4(a), the gaze of a subject follows the edges of a rectangular box. It can be seen that 3D gaze points are indeed obtained along the edges of the box. Finally, in Fig. 4(b), landmarks are arranged in a spiral-like manner. On the basis of the above results, we can say that the proposed inside-out camera can be used to estimate 3D gaze points and to obtain information on where in 3D space someone is looking, which should be useful when attempting to understand a person's behavioral intentions.

5. Conclusion

We proposed an "inside-out camera" for realizing First Person Vision and a technique for estimating a person's gaze point using the structural features of the camera. The inside-out camera uses half-mirrors to enable a person's eyeball to be captured directly from the front and the person's visual field to be captured from a position nearly equivalent to the eyeball. The gaze point is calculated from a gaze vector using conversion-equation parameters computed beforehand. It was found from an evaluation experiment that gaze point could be estimated with an average error of about 15 pixels.

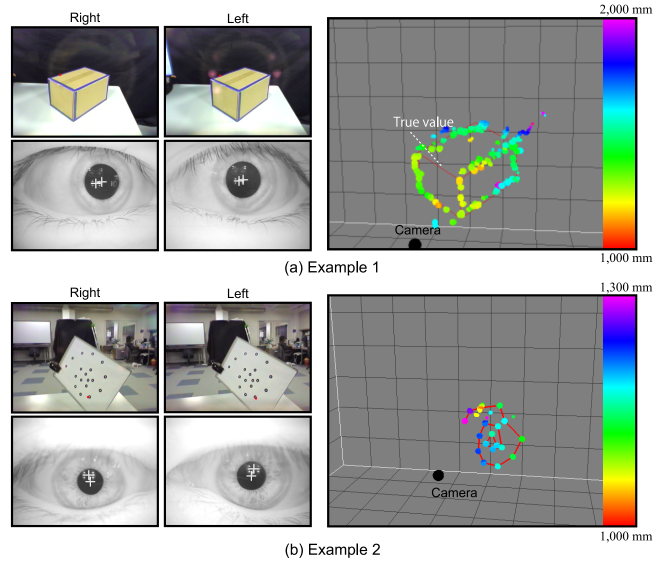

Figure 4. Estimation of 3D position of gaze points

References

[1] T. Kanade. First-person, inside-out vision. In *IEEE Workshop on Egocentric Vision, CVPR*, 2009. 1

[2] K. Matsuda, T. Nagami, and S. Yamane. A system for measuring eye position. *Technical report of IEICE. Thought and language*, 100(47):9–16, 2000. 2

[3] T. Ohno. Quick menu selection task with eye mark. *Transactions of Information Processing Society of Japan*, 40(2):602–612, 1999. 3

[4] Y. Sakashita, H. Fujiyoshi, Y. Hirata, and N. Fukaya. Real-time measurement system of cyclodction movement based on fast ellipse detection. *IEEJ Transactions on Electrical and Electronic Engineering*, 127-C:591–598, 2007. 2

[5] R. Stiefelhagen, J. Yang, and A. Waibel. Tracking eyes and monitoring eye gaze. In *Proceedings of the Workshop on Perceptual User Interfaces*, pages 98–100, 1997. 2

[6] T. Takegami and T. Gotoh. Vision-based algorithm for eye movement detection using corneal reflection and iris contour. *The transactions of the Institute of Electronics, Information and Communication Engineers. D-I*, 82(10):1295–1303, 1999. 2

[7] H. Tanaka, S. Hikita, T. Kasai, and T. Takeda. Eye-gaze detection based on the position of the cornea curvature center using two near-infrared light sources. *IEICE technical report. ME and bio cybernetics*, 108(479):177–180, 2009. 2