Acquisition of 3D Gaze Information from Eyeball Movements using Inside-out Camera

Shoichi Shimizu
Advanced Technology R&D Center, Mitsubishi Electric Corporation
8-1-1, Tsukaguchi-Honmachi, Amagasaki, Hyogo 661-8661, Japan
Shimizu.Shoichi@ab.MistubishiElectric.co.jp

Hironobu Fujiyoshi
Dept. of Computer Science, Chubu University
1200 Matsumoto, Kasugai, Aichi 487-8501
Japan
hf@cs.chubu.ac.jp

ABSTRACT

We propose a method for obtaining 3D gaze information using inside-out camera. Such information on 3D gaze points can be useful not only to clarify higher cognitive processes in humans but also to reproduce the 3D shape of an object from eyeball movement simply by gazing at the object as an extension of the visual function. Using half-mirrors, an inside-out camera can capture a person's eyeball head-on and can capture the person's visual field from a position equivalent to that of the eyeball. Here, the relationship between the gaze vector obtained from images of the eyeball and the gaze point in images capturing the visual field is expressed by a conversion equation. The 3D position of the gaze point can then be estimated by using stereo constraints in two scene cameras. In an evaluation experiment, the gaze point could be estimated with an average error of about 15 pixels, and we also showed the 3D scan path obtained by the proposed method from eyeball movement by gazing at the object.

Categories and Subject Descriptors

I.2.10 [**Vision and Scene Understanding**]: 3D/stereo scene analysis

General Terms

MEASUREMENT

Keywords

3D gaze point, inside-out camera, stereo vision

1. INTRODUCTION

The human eye enables a person to instantly absorb information and act accordingly. As a consequence, information on a person's gaze, which reveals what objects in the outside the world the person is looking at, can be valuable in determining that person's behavioral intentions. Noton et al. discovered that similar scanpaths are used when a person is shown the same object at different times[1]. Obtaining information on eye movement in this way can therefore be expected to clarify higher cognitive processes in humans [2].

Camera systems for detecting a person's gaze have been commercially available for some time. These systems have been used for rehabilitation purposes in relation to impaired visual perception functions, for comparing and evaluating the visual perception of road signs, and for evaluating product usability [3, 4, 5, 6, 7]. Additionally, the Aided Eyes system proposed by Ishiguro et al. uses information on eye activity obtained by a sensor to augment human memory [8]. These commercialized gaze measurement equipment consist of a camera for capturing the subject's eyeball and a scene camera for capturing the subject's visual field. Due to structural considerations in camera installation, however, a scene camera cannot be placed at the same position as the eyeball, which means that the image obtained of the visual field is not the same as the actual visual field. Another problem is that the eyeball cannot be captured head-on when attempting to measure gaze from eyeball images since that would obstruct the subject's visual field.

In view of the above problems, we propose a method for obtaining 3D gaze information using inside-out camera. Using half-mirrors, an inside-out camera can capture a person's eyeball head-on and can capture the person's visual field from a position equivalent to that of the eyeball. Here, the relationship between the gaze vector obtained from images of the eyeball and the gaze point in images capturing the visual field is expressed by a conversion equation. The 3D position of the gaze point can then be estimated by using stereo constraints in two scene cameras. Such information on 3D gaze points can be useful not only to clarify higher cognitive processes in humans but also to reproduce the 3D shape of an object from eyeball movement simply by gazing at the object as an extension of the visual function.

2. EXISTING GAZE MEASUREMENT EQUIPMENT

In this section, we describe two types of existing gaze measurement equipment-head-mounted and standalone-and point out the advantages and disadvantages of each.

2.1 Head-mounted type

The head-mounted type of gaze measurement equipment [3, 4, 5] can capture both of the subject's eyes and the subject's visual field, using two cameras for capturing the eyeballs and one camera for capturing the visual field. This type of equipment can be used to analyze the eye gaze of athletes, and it can be used in computer games by attaching a head-mounted display.

2.2 Standalone type

The standalone type of gaze measurement equipment [6, 7] does not make contact with the subject enabling eye gaze to be measured in a natural state. It cannot, however, capture an image of the sub-

Table 1: Comparison of both types

	Head-mounted	Standalone
Subject movement	Unrestricted	Restricted
Burden on subject	Medium	Low
Accuracy	High	Medium
Visual-field capture	Yes	No

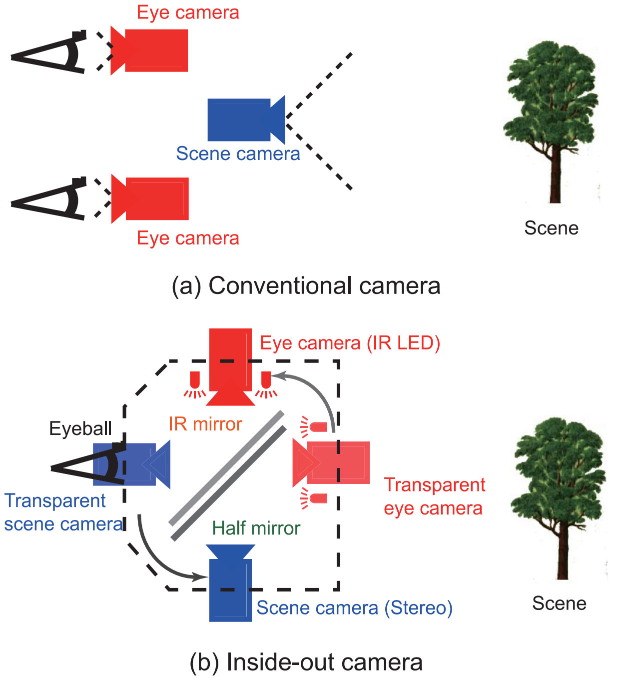

Figure 1: Comparing conventional camera and inside-out camera

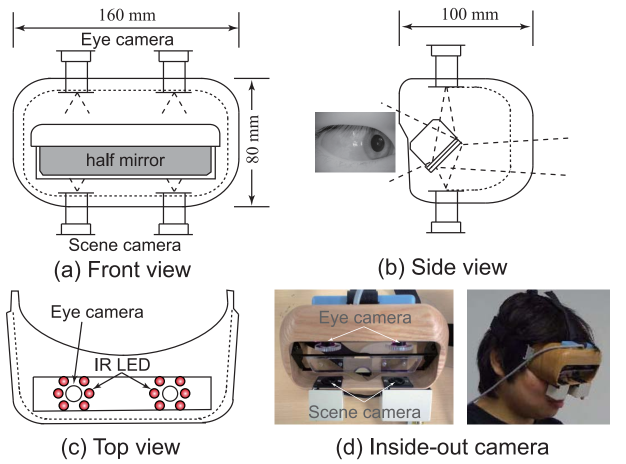

Figure 2: Inside-out camera system

ject's visual field. This type of equipment imposes little burden on the subject and can be easily used to measure the eye gaze of either children or adults.

2.3 Comparison of head-mounted and standalone types

The advantages and disadvantages of the head-mounted and standalone types of gaze measurement equipment are listed in Table 1. As the name implies, the head-mounted type of equipment is worn by the subject enabling freedom of movement. This type of equipment can be used, for example, to analyze the eye gaze of athletes. On the other hand, equipment attached to the human body can be a burden to the subject, which means that a subject's gaze when being measured might be different from the subject's gaze in everyday situations. The standalone type of equipment, meanwhile, cannot capture the subject's visual field; in terms of subject behavior, it can only capture the subject's eyeballs. Its advantage, however, is that it can measure gaze without the subject being aware of being measured, enabling data to be obtained under natural conditions.

2.4 Problems with existing gaze measurement equipment

As shown in Fig. 1(a), the scene camera for capturing the subject's visual field in existing gaze measurement equipment is placed at a position apart from the eyeballs, which means that parallax will occur between the subject's actual visual field and the captured image. Thus, as in the case of multi-camera stereo vision, parallax between the subject's visual field and the camera image will increase as the distance to the object being captured decreases. As a result, it has been common practice when using existing gaze measurement equipment to set some sort of constraints on the subject's gaze point. For example, the target of the subject's gaze may have to be set at a certain distance, that is, it may have to lie on a known plane.

3. INSIDE-OUT CAMERA

We here describe our proposed inside-out camera.

3.1 Equipment configuration

Our most recent prototype equipment has the shape of goggles as shown in Fig. 2. 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. 1(b), 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.

3.2 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. Examples of images captured with the eye cameras are shown in Fig. 3. The LEDs arranged around each camera emit near-infrared light in the 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.

3.3 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

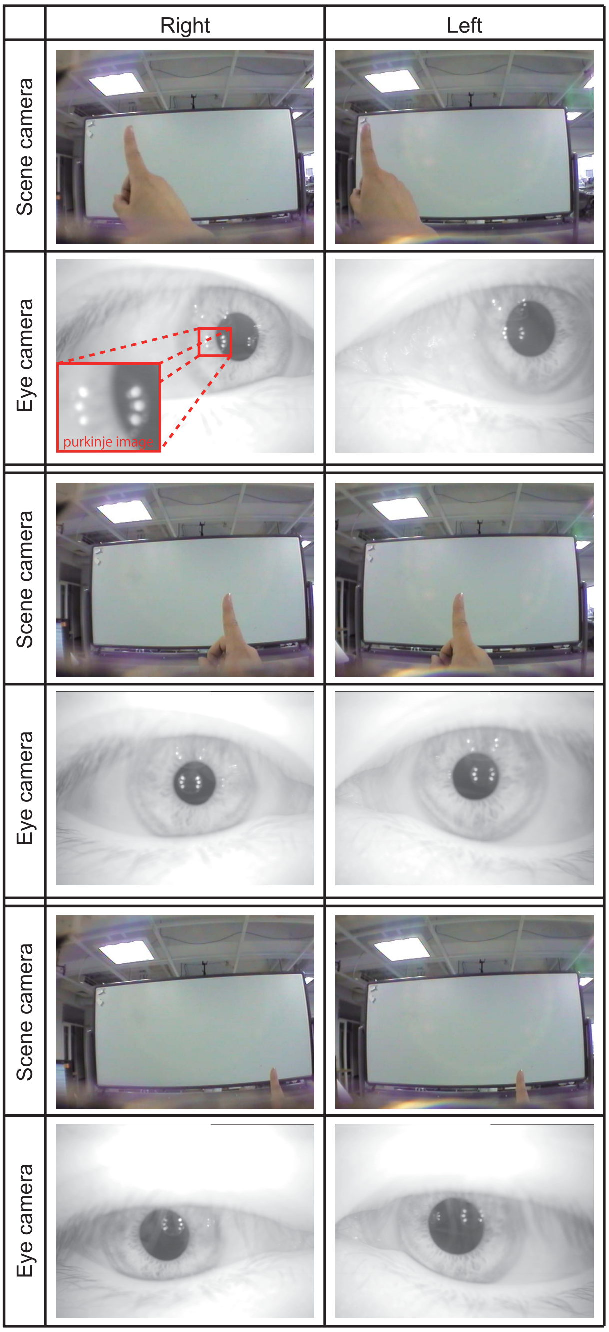

Figure 3: Images captured with inside-out camera

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 [9] or Zhang model [10] and to estimate the 3D position of the gaze point in the visual field. Examples of images captured with these scene cameras are also shown in Fig. 3.

3.4 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.

4. 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 [11, 12, 13, 14]. Since the gaze vector moves from one landmark to another, a correlation clearly exists between the gaze vector and landmarks being gazed at. With the proposed inside-out camera, this correlation is easy to work with as described in section 3.4. As shown by images of Fig. 3, if the gaze point is moved to the left, the gaze vector estimated from the Purkinje images and pupil center also moves to the left, and if the gaze point is moved to the right, the gaze vector also moves to the right. 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
   Estimate the gaze vector from the pupil center and Purkinje images[1]

2. Calculate gaze point
   Convert gaze vector to gaze point using a conversion equation

3. Calculate 3D gaze point
   Use the gaze points of both eyes to calculate the 3D gaze point in stereo vision

4.1 Estimation of gaze vector

The gaze detection method proposed by Ohno et al. calculates the cornea curvature center from a Purkinje image and takes the line connecting this point and the pupil center to be the gaze vector [15]. It has been reported that this technique can estimate the gaze vector with better accuracy than the technique using the eyeball rotation center and pupil center. Our proposed technique also estimates the gaze vector from the cornea curvature center and pupil center.

First, the cornea curvature center is calculated from Purkinje images with the aim of estimating the gaze vector. In the method of Ohno et al. [15], there is only one light source, and the cornea curvature center is calculated by correcting the position of the Purkinje image so that it falls on the camera's optical axis.

In contrast, 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. Purkinje images can be extracted from the techniques proposed in Refs. [15, 14], and the cornea curvature center $\boldsymbol{C} = [C_u, C_v]^T$ on the video image

[1]Purkinje images are reflected lights of LEDs in the surface of the eye.

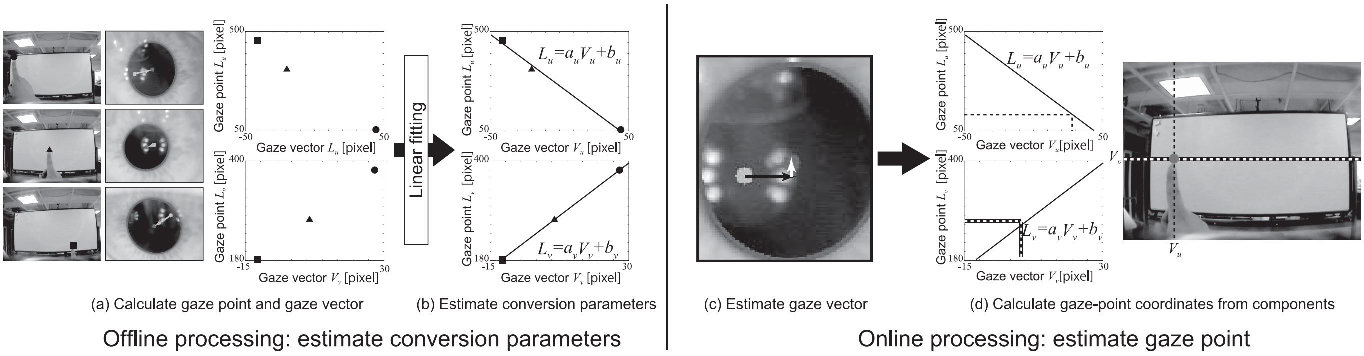

Figure 4: Relation between gaze point and gaze vector

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 $\boldsymbol{P}$ [16].

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:

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

4.2 Calculation of 2D gaze point using conversion equation

The gaze point on the scene video can be calculated using the gaze vector estimated in section 4.1 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. 4 and described below.

4.2.1 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. 4(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. 4(b), gaze point L can be calculated on input of gaze vector $\boldsymbol{V}$.

4.2.2 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 4.1. 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. 4(d).

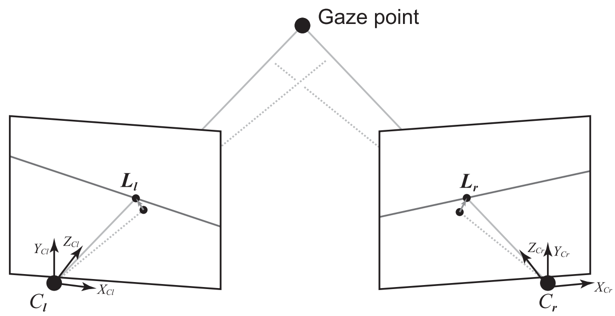

Figure 5: Optimal correction of gaze points

4.3 Optimal correction of gaze points

The gaze points for both eyes calculated in section 4.2 includes error in the estimated gaze vectors producing an offset between these gaze points and, as a result, an offset between the gaze points on the scene videos. We therefore correct the positions of the two gaze points on the scene videos so that they coincide using the optimal correction technique proposed in Ref.[17] as shown in Fig. 5.

Given that point $\boldsymbol{L}_l$ in the scene video on the left corresponds to point $\boldsymbol{L}_r$ in the scene video on the right, the lines of sight that pass through each of these points will intersect under the condition that Eq.(4) is satisfied, where each point is expressed in terms of homogeneous coordinates $m_l = [L_{ul}/f_0, L_{vl}/f_0, 1]^T$ and $m_r = [L_{ur}/f_0, L_{vr}/f_0, 1]^T$.

$$\boldsymbol{m}_r^T \boldsymbol{F} \boldsymbol{m}_l = 0 \tag{4}$$

Here, $\boldsymbol{F}$ denotes a fundamental matrix that indicates the geometrical relationship between left and right images.

In optimal correction, gaze points on the scene videos are moved so that the corresponding lines of sight intersect but in a manner that minimizes the amount of this movement. Left-side and right-side scene-video coordinates $\hat{\boldsymbol{m}}_l$ and $\hat{\boldsymbol{m}}_r$ after optimal correction of the left-side and right-side scene cameras are calculated by Eqs.(5) and (6), respectively.

$$\hat{\boldsymbol{m}}_l = \boldsymbol{m}_l - \frac{(\boldsymbol{u}\boldsymbol{\xi})\boldsymbol{P}_k \boldsymbol{F} \boldsymbol{m}_r}{(\boldsymbol{u}\boldsymbol{V}_0[\xi]\boldsymbol{u})} \tag{5}$$

$$\hat{\boldsymbol{m}}_r = \boldsymbol{m}_r - \frac{(\boldsymbol{u}\boldsymbol{\xi})\boldsymbol{P}_k \boldsymbol{F}^T \boldsymbol{m}_l}{(\boldsymbol{u}\boldsymbol{V}_0[\xi]\boldsymbol{u})} \tag{6}$$

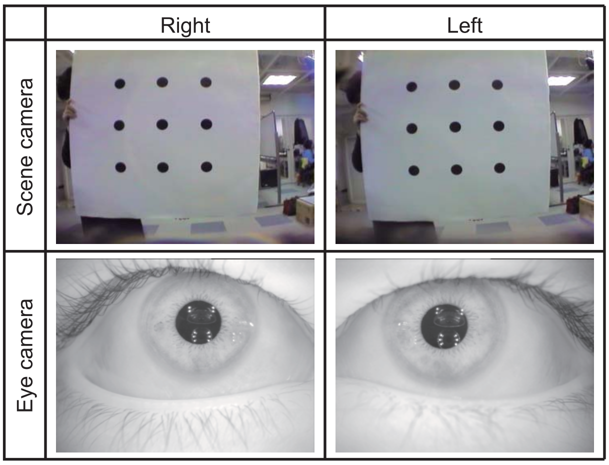

Figure 6: Example of experimental data

Here, $\boldsymbol{P}_k$ is a matrix whose third component is set to 0.

$$\boldsymbol{P}_k = \mathrm{diag}(1,1,0) \tag{7}$$

Additionally, $\boldsymbol{u}$ and $\boldsymbol{\xi}$ rewrite fundamental matrix $\boldsymbol{F}$ and data $\boldsymbol{m}$, respectively, into nine-dimensional vectors.

$$\begin{aligned} \boldsymbol{u} &= [F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33}]^T \\ \boldsymbol{\xi} &= [L_{u_l}L_{u_r}, L_{u_l}L_{v_r}, f_0L_{u_l}, L_{v_l}L_{u_r}, \\ &\quad L_{v_l}L_{v_r}, f_0L_{v_l}, f_0L_{u_r}, f_0L_{v_r}, f_0^2]^T \end{aligned}$$

Here, matrix $\boldsymbol{V}_0[\boldsymbol{\xi}]$ is a 9×9 matrix defined in Ref.[17]. Calculated $\hat{\boldsymbol{m}}$ is substituted in m and Eqs.(5) and (6) are recalculated. This process is repeated until convergence is achieved.

4.4 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 matching. We can solve the problem of the gaze vectors not intersecting in the 3D space by treating it as a problem in stereo matching. Given that the human viewing angle during gazing is about two degrees [18], we calculate a 3D position from the regions of both gaze points. Since the 3D gaze point is calculated by using stereo matching, its accuracy of 3D gaze point is equal to that of stereo matching.

5. EVALUATION EXPERIMENT

To evaluate the proposed technique, we collected data on actual landmark gazing and calculated the error between the true value of a landmark and the estimated gaze point. This section presents the results of this experiment and presents examples of estimating a 3D gaze point.

5.1 Experimental setup

Data was collected as follows. As shown in Fig. 6, a board with nine landmarks was prepared and gazing was performed in order from the top left for board depths from 1 to 5m at intervals of 50 cm. Data for 54 points were collected at each interval and a total of 486 sets of scene video and eyeball video were obtained. Five subjects participated in this experiment.

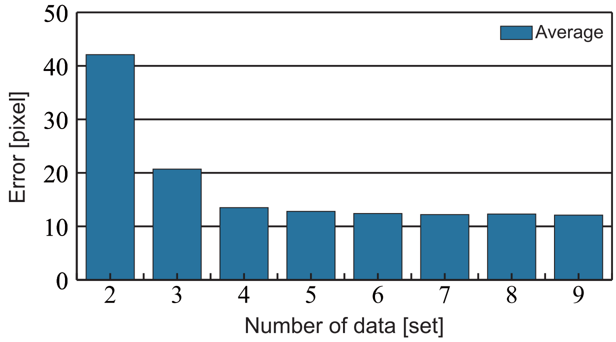

Figure 7: Error versus number of data used for calibration

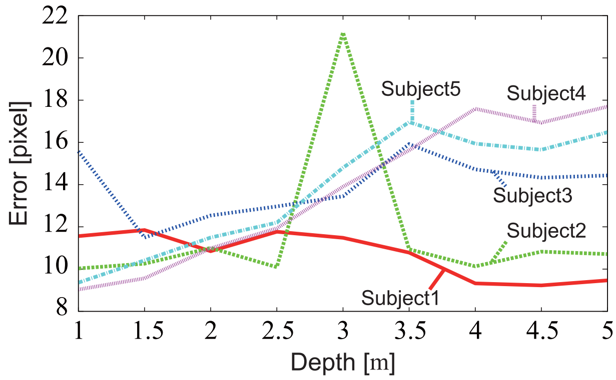

Figure 8: Error versus depth of 3D gaze point

5.2 Error estimation results

The proposed technique requires that parameters for the conversion equation be calculated beforehand. Needless to say, the accuracy of these parameters depends on calibration data. We therefore investigated error in results when varying the amount of data used for calibration. Specifically, we randomly selected from two to nine sets of data from the 54 sets of data obtained at a depth of 1 m and performed 100 trials for each collection of data sets. Human viewing angle during gazing has been reported to be about two degrees [18], and with this in mind, we took error range to be a diameter of about 20 pixels when projecting a two-degree region on scene video.

Figure 7 shows average error for 100 trials versus number of calibration data sets. These results indicate that an average error of less than 15 pixels can be obtained when using four or more sets of calibration data and that error can be kept within the error range of 20 pixels. It can also be seen that average error tends to decrease as the number of data sets increases indicating that gaze-point estimation can be stabilized by increasing the number of data sets used for calibration.

Finally, we investigated the average error of each depth. The average error of using 9 calibration data sets versus the depth of 3D gaze point is shown in Figure 8. The average error of the subject4 and that of subject5 increased commensurately with the depth. However, the average error of the remaining three subjects did not increased commensurately with the depth. Therefore, we can see that the estimation of 2D gaze point is not easily affected by the distance of the 3D one.

5.3 Estimation of 3D gaze point

Figure 9 shows three examples of estimating 3D gaze points from calculated gaze points. In Fig. 9(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. Next, in Fig. 9(b), the subject changes gaze target in the order of ①sign, ② human being, and ③ box. The corresponding 3D gaze points could be obtained, and the general position of the subject's gaze could be understood and a 3D ScanPath obtained. Finally, in Fig. 9(c), landmarks are arranged in a spiral-like manner. As in the example of Fig. 9(b), a 3D ScanPath of gaze points could be obtained. 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.

6. CONCLUSION

We proposed a method for acquiring of 3D gaze information from eyeball movements using the "Inside-out Camera" 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. However, we think that we need to improve the usability of the system, and the inside-out camera is just a prototype. In particular, we will have to reduce the weight of the camera and to make it more comfortable to wear.

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. We also showed the 3D scan path obtained by the proposed method from eyeball movement by gazing at the object.

7. REFERENCES

[1] David Noton and Lawrence Stark. Eye movements and visual perception. *Scientific American*, 224:35–43, 1971.

[2] Takehiko Ohno. What can be learned from eye movement? : Understanding higher cognitive processes from eye movement analysis. *Cognitive studies : bulletin of the Japanese Cognitive Science Society*, 9(4):565–579, 2002.

[3] NAC Image Technology Inc. eyemark recorderEMR-9.

[4] Arrington Research, inc. EyetrackerViewPoint.

[5] ISCAN. AA-ETL-500B.

[6] NAC Image Technology Inc. eyemark recorderEMR-AT VOXER.

[7] Tobii. Tobii XL.

[8] Yoshio Ishiguro, Adiyan Mujibiya, Takashi Miyaki, and Jun Rekimoto. Aided eyes: Eye activity sensing for daily life. In *The 1st Augmented Human International Conference (AH2010)*, 2010.

[9] Roger Y. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323–344, 1922.

[10] Z. Zhang. A flexible new technique for camera calibration. *PAMI*, 22(11):1330–1334, 2000.

[11] Rainer Stiefelhagen, Jie Yang, and Alex Waibel. Tracking eyes and monitoring eye gaze. In *Proceedings of the Workshop on Perceptual User Interfaces*, pages 98–100, 1997.

[12] Takeshi Takegami and Toshiyuki 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.

[13] Keiji Matsuda, Takeshi Nagami, and Shigeru Yamane. A system for measuring eye position. *Technical report of IEICE. Thought and language*, 100(47):9–16, 2000.

[14] Hiroaki Tanaka, Shinichi Hikita, Takeshi Kasai, and Takashi 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.

[15] Takehiko Ohno, Naoki Mukawa, and Atsushi Yoshikawa. Freegaze: a gaze tracking system for everyday gaze interaction. In *the symposium on ETRA 2002: eye tracking research & applications symposium*, pages 125–132, 2002.

[16] Yusuke Sakashita, Hironobu Fujiyoshi, Yutaka Hirata, and Naoki 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.

[17] Kenichi Kanatani, Yasuyuki Sugaya, and Hirotaka Niitsuma. Triangulation from two views revisited: Hartley-sturm vs. optimal correction. *Proceedings of the 19th British Machine Vision Conference (BMVC'08)*, 8:173–182, 2008.

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

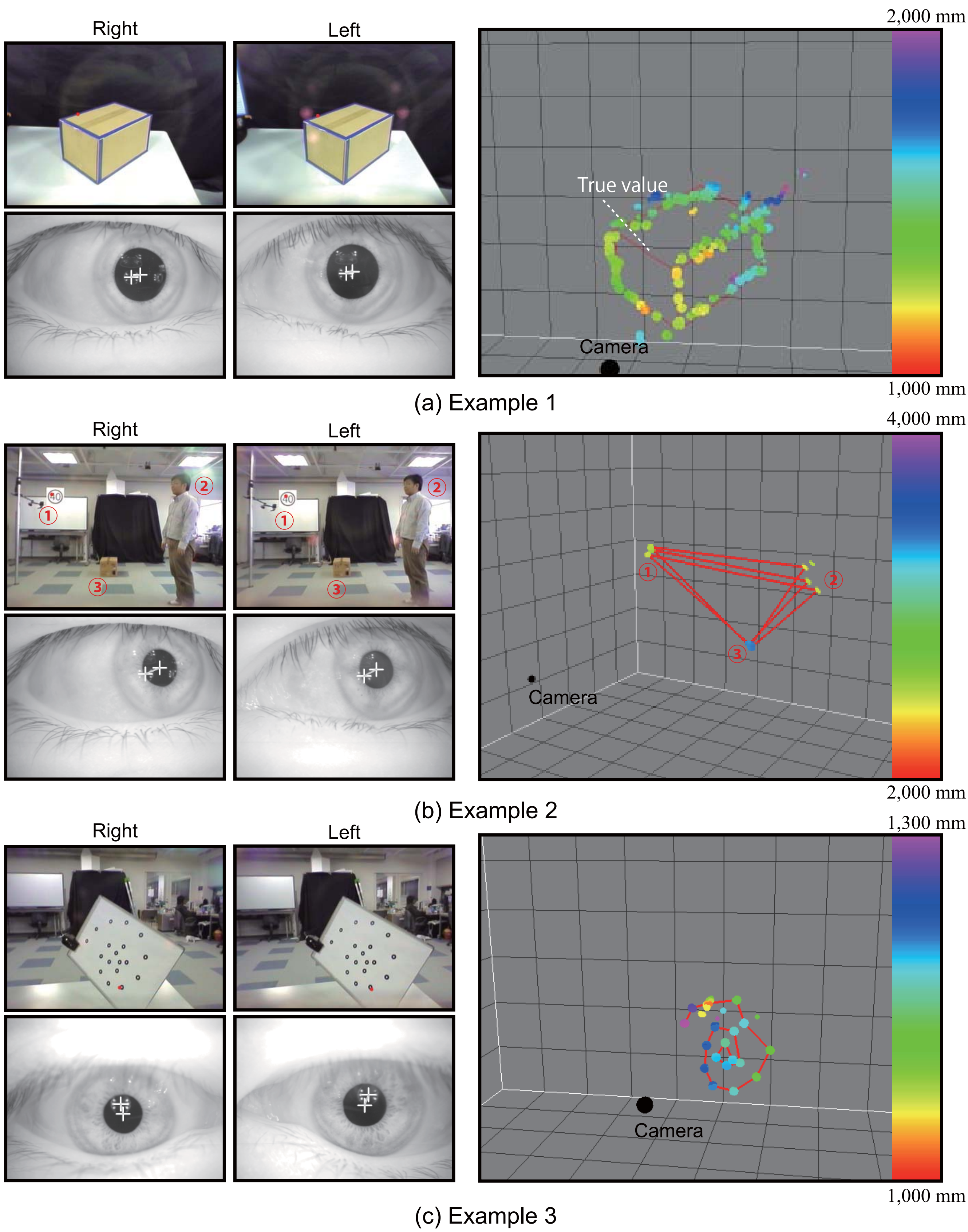

Figure 9: Estimation of 3D position of gaze points