Head-Free, Remote Gaze Detection System Based on Pupil-Corneal Reflection Method with Using Two Video Cameras – One-Point and Nonlinear Calibrations

Yoshinobu Ebisawa and Kiyotaka Fukumoto

Graduate School of Engineering, Shizuoka University, Hamamatsu, 432-8561 Japan
ebisawa@sys.eng.shizuoka.ac.jp

Abstract. We developed a pupil-corneal reflection method-based gaze detection system, which allows head movements and achieves easy gaze calibration. The proposed gaze detection theory determines gaze points on a PC screen from the vector from the corneal reflection to pupil center, 3D pupil position, two cameras position, etc. In a gaze calibration procedure, after a user is asked to gaze one specific calibration target at a center of a PC screen, the nonlinear characteristic of the eyes has been automatically corrected while the user is using this gaze system. The experimental results show that the proposed calibration method improved the precision of gaze detection during browsing web pages. In addition, the average gaze error in the visual angle is less than 0.6 degree for the nine head positions.

Keywords: Gaze detection, Gaze calibration, Head movement, Pupil.

1 Introduction

For human interface and human behavior monitoring, precise eye-gaze detection with easy calibration procedure is desired. Current commercial gaze detection systems need the gaze more than five points on a PC screen for high precision. However, when the gaze detection system is used for the general public or infants, it is difficult that a user is asked to gaze some points. In previous studies, the calibration methods to gaze a few or no gaze points have been proposed [1][2]. However, their methods include some problems such as a range of user's head movement and easiness for field surveys. Thus, in our previous study, we have developed a gaze detection system based on the pupil-corneal reflection method, which allows large head movements and achieves easy gaze calibration [3].

In this system, an optical system for detecting the pupil and corneal reflection images consist of a camera and a two concentric near-infrared LED ring light source attached to the camera. The inner and outer LED rings generate bright and dark pupil images, respectively. The pupils are detected from the difference image created by subtracting the bright and dark pupil images. Fig. 1 shows the gaze detection theory in

M. Kurosu (Ed.): Human-Computer Interaction, Part IV, HCII 2013, LNCS 8007, pp. 205–214, 2013.

3D space. The 3D coordinates of the pupils are determined by the stereo matching method using two optical systems. The vector from the corneal reflection center to the pupil center in the camera image is replaced by its actual size vector $\boldsymbol{r}$. The angle between the line of sight and the line passing through the pupil center and the camera (light source) is denoted as θ. The relationship is assumed as $\theta = k|\boldsymbol{r}|$ where k is a constant. The theory allowed head movement of the user and facilitates the gaze calibration procedure. In the automatic calibration method, calibration procedure is accomplished while the user looks around on the PC screen without fixating on any specific calibration target. In the one-point calibration method, the user is asked to fixate on one calibration target at the center of the PC screen in order to correct the innate difference, ΔQ, between the optical and visual axes of the eyeball. In the *two-point calibration* method, in order to correct the nonlinear relationship between θ and $|\boldsymbol{r}|$, which occurs where θ is large, the user is asked to fixate on another target presented at the top of the PC screen as well as the center target. The experimental results show that the three proposed calibration methods improve the precision of gaze detection step by step. In addition, the average gaze error in the visual angle is less than 1 degree for the seven head positions of the user.

However, afterwards, we found a simpler calibration method. In the method, nonlinear relationship between k and $\boldsymbol{r}$ is automatically corrected during the user look around the PC screen area after the one-calibration method is completed. In the present paper, we propose the new calibration and gaze detection methods. The method simultaneously deals with the two problems of the difference, ΔQ, between the optical and visual axes and the nonlinear relationship between θ and $|\boldsymbol{r}|$. In addition, the proposed gaze detection theory became more simple and flexible than our previous theory.

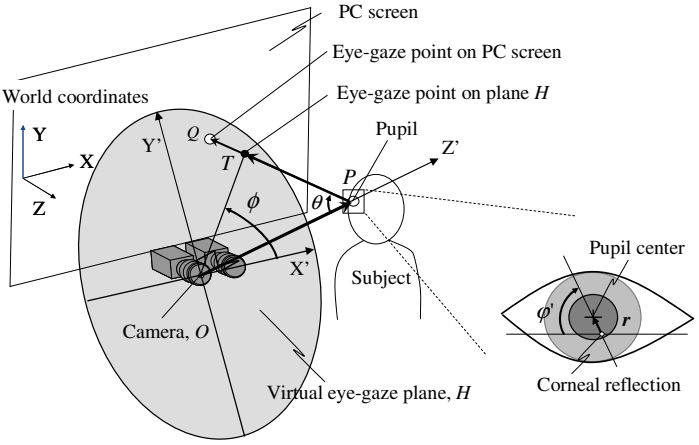


Fig. 1. Gaze detection theory in 3D space

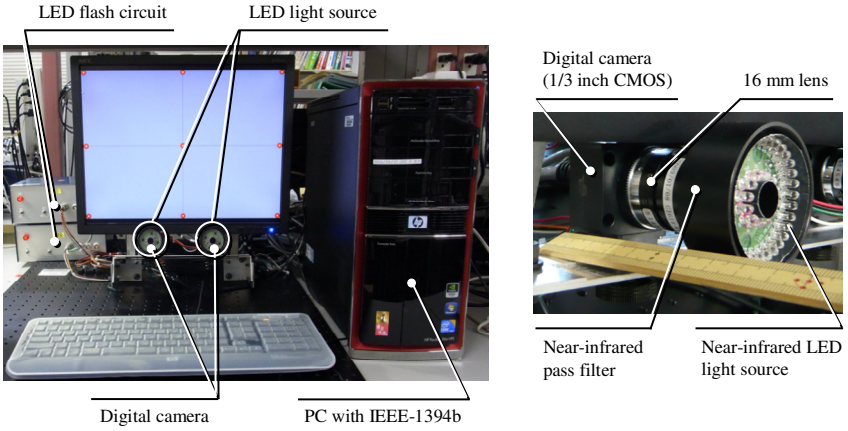
2 Gaze Detection System

2.1 Pupil Detection Principle and Method

In our previous study [3], a light source consisting of near-infrared LEDs that are arranged in two concentric rings (inner and outer rings) was proposed in order to create the bright and dark pupil images, respectively. Since the inner ring is located near the aperture of the camera, the inner ring generates a bright pupil image. Since the outer ring is far from the aperture, the outer ring generates a dark pupil image. In addition, in order to miniaturize the light source, we used the LEDs having two different wavelengths. The structure of the light source is described in the next section.

2.2 System Configuration

Fig. 2 (a) shows an overview of the developed gaze detection system. This system has two optical systems (Fig. 2 (b)), each of which consists of a digital video camera having near-infrared sensitivity, a 16-mm lens, an infrared filter (IR80), and a light source. Light sources consisting of near-infrared LEDs of two different wavelengths that are arranged in two concentric rings (inner: 850 nm, outer: 940nm) are attached to each camera. The pupil becomes brighter in the 850nm ring than the 940nm ring because the transmissivity of the eyeball medium is different. The distance between the LEDs and the aperture of the camera also varies the pupil brightness. The combined effects of the distance and the difference in transmissivity were applied to the light source. An internal synchronization at the hardware level is possible if the cameras are connected to buses of the IEEE-1394 PCI board. The internal synchronization was used to drive two cameras with a slight synchronization difference (670μs) because of avoiding mutual light interference of the optical systems.



(a) Overview of the gaze detection system (b) The optical system

Fig. 2. Overview of the gaze detection system and the optical system

An 8-bit gray scale image of a user's face was input into a personal computer (PC) via the board. The captured image sizes were 640 x 480 pixels. The image was processed using the PC to detect the centers of the pupils and the corneal reflections, which were used to determine the gaze points.

2.3 Image Processing for Detection of the Centers of Pupil and the Corneal Reflection Image

The pupil is detected from the difference image generated from the bright and dark pupil images (Fig. 3 (a)-(c)). The image is processed in the following order: binarization, removal of isolated pixels, noise reduction using mathematical morphology operations, and labeling. The largest and second largest labeled regions were detected as the two pupils. When a pupil was detected in prior frames, the pupil position in the current frame was estimated using a linear Kalman filter, and a small window (70 x 70 pixels) was then applied around the estimated pupil position. On the other hand, when the pupil was not detected in the prior frames (e.g., effect of blinks and eyelashes), the pupils were again searched for in the entire image of the user's face.

The image within the small window is transformed into an image with twice the resolution (140 x 140 pixels) (Fig. 3 (d)). This image from the bright and dark pupil images is processed by binarization and labeling, and an intense and tiny label closest to a center of the double-resolution image is determined as the corneal reflection. The pupil region was again determined by binarizing the difference image, which was obtained after shifting this double-resolution dark pupil image so that the corneal reflection in this dark pupil image may coincide with that in the double-resolution bright pupil image [4]. This process helped to decrease the positional deviation between the bright and dark pupil images while the user's head is moving. Ellipse fitting for the contour of the pupil was then performed. The center of the ellipse was determined as the center of the pupil. The center of gravity considering the values of the pixels in the corneal reflection region in the bright pupil image was determined as the center of the corneal reflection.

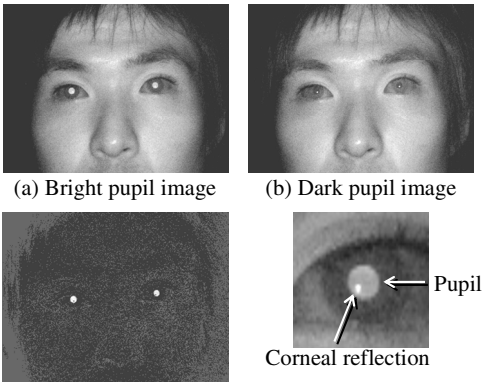


Fig. 3. Detection of pupils and corneal reflection

2.4 Gaze Detection Theory and Calibration Method

In the Fig. 4, O_1' and O_2' indicate the pinholes of two stereo-calibrated cameras. We assume that the light source attached to each camera is located at the same position as the corresponding camera. The 3D pupil position P is obtained by the stereo-matching method. The optical axis of an eyeball passes through the pupil P and gaze point Q on the screen plane of the PC display. Now we define the virtual gaze planes (H_1 and H_2) of the cameras for one eyeball. These planes are vertical to the line passing through P and O_1' or O_2', and they pass through O_1' and O_2'. The X-axis (X_1 or X_2) of planes H_1 and H_2 is determined as the line intersection between the corresponding plane and the horizontal plane in the world coordinate system (x − y − z). H_1 and H_2 rotate according to the displacements of the pupil in the world coordinate system.

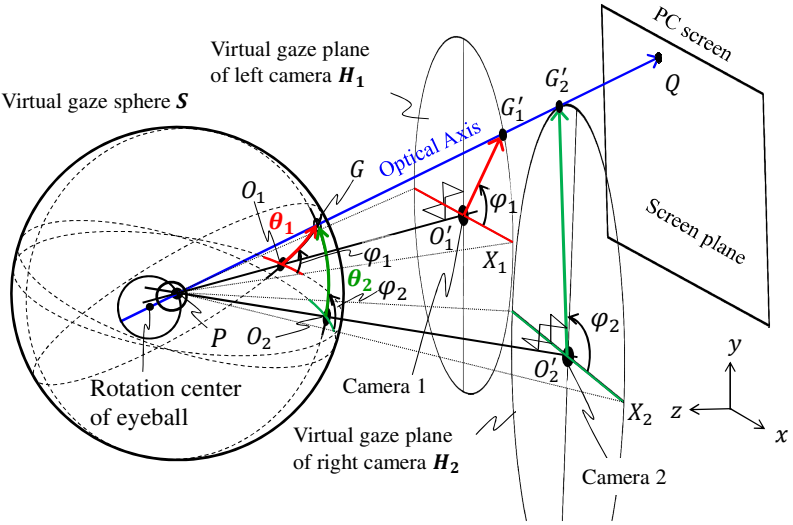


Fig. 4. Gaze detection theory using visual gaze sphere

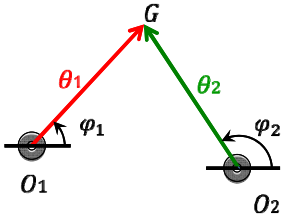


Fig. 5. Angular vectors transformed into 2D coordinate system

Next, we define the virtual gaze sphere S whose center is P. The optical axis PQ has intersection points with sphere S, H_1 and H_2. The intersection points are denoted as G, $G_1'$and G_2', respectively. Angular vectors $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$ on sphere S can be defined

as projections of vectors $O_1'G_1'$ and $O_2'G_2'$ to sphere S. In addition, horizontal axes X_1 and X_2 of planes H_1 and H_2 are projected to sphere S. Here, orientations φ_1 and φ_2 of vectors $O_1'G_1'$ and $O_2'G_2'$ are also projected to sphere S. According to these projections, angular vectors $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$ on sphere S are transformed into the 2D vectors on a flat plane as shown in **Fig. 4**. Here, vector $\overrightarrow{O_1O_2}$ is expressed as follows:

$$\overrightarrow{O_1O_2} = \boldsymbol{\theta_1} - \boldsymbol{\theta_2} \tag{1}$$

Since you can see both relationships $|\boldsymbol{\theta_1}| = \angle O_1PG$ and $|\boldsymbol{\theta_2}| = \angle O_2PG$ referring to Fig. 1, the following equation is obtained.

$$\left|\overrightarrow{O_1O_2}\right| = \angle O_1PO_2 \tag{2}$$

Here, we assume that the angular vector $\boldsymbol{\theta}$ is a function of $\boldsymbol{r}$ as

$$\boldsymbol{\theta} = f(\boldsymbol{r}). \tag{3}$$

As mentioned before, $\boldsymbol{\theta}$ is an angular vector having size θ and orientation φ, and f is monotonically increasing function. In one-point calibration, we assume a linear function as follows:

$$\boldsymbol{\theta} = k\boldsymbol{r} \tag{4}$$

where k is a constant. In general, there is a difference between the optical axis and visual axis. In order to compensate it, measured vector $\boldsymbol{r}'$ is compensated by offset vector $\boldsymbol{r_0}$.

$$\boldsymbol{r} = \boldsymbol{r}' - \boldsymbol{r_0} \tag{5}$$

Accordingly, the following equations are obtained for cameras 1 and 2 from equations (4) and (5).

$$\boldsymbol{\theta_1} = k\boldsymbol{r_1} = k(\boldsymbol{r_1'} - \boldsymbol{r_0}) \tag{6}$$

$$\boldsymbol{\theta_2} = k\boldsymbol{r_2} = k(\boldsymbol{r_2'} - \boldsymbol{r_0}) \tag{7}$$

Here, $\boldsymbol{r_1'}$ and $\boldsymbol{r_2'}$ are the pupil-corneal reflection vectors measured from cameras 1 and 2, respectively. $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$ are the compensated vectors. From equation (1), (2), (6) and (7), k is given by the following equation.

$$k = \left|\frac{\boldsymbol{\theta_1} - \boldsymbol{\theta_2}}{\boldsymbol{r_1'} - \boldsymbol{r_2'}}\right| = \frac{\boldsymbol{\angle O_1PO_2}}{|\boldsymbol{r_1'} - \boldsymbol{r_2'}|} \tag{8}$$

Using this value of k, $\boldsymbol{r_0}$ is determined from equations (6) and (7). Here, $\boldsymbol{r_0}$ is common for both cameras. Determining the values of k and $\boldsymbol{r_0}$ means gaze calibration in the *one-point calibration* method. In the *one-point calibration* procedure, the calibrations parameters are determined when a subject fixates on a visual target presented at the center of the PC screen. In the *gaze detection procedure*, first, the pupil-corneal reflection vectors $\boldsymbol{r_1'}$ and $\boldsymbol{r_2'}$ are obtained from the images of the two

cameras. By using equations (6) and (7), $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$ are calculated. Next, the corresponding visual axes are determined from $\boldsymbol{\theta_1}$, $\boldsymbol{\theta_2}$ and pupil position $\boldsymbol{P}$. Finally, the gaze points on the screen are estimated as the intersection points between the screen plane and the visual axis.

In order to compensate the *nonlinear* relationship between θ and $|\boldsymbol{r}|$, the following equation were used.

$$\theta = f(\boldsymbol{r}) = g(\boldsymbol{r})|\boldsymbol{r}| \tag{9}$$

where $g(\boldsymbol{r}) = h|\boldsymbol{r}'|^2 + k_0 \equiv k_f$. Therefore, θ is denoted and calculated by the following equation.

$$\theta = k_f|\boldsymbol{r}| = (h|\boldsymbol{r}'|^2 + k_0)|\boldsymbol{r}' - \boldsymbol{r_0}| \tag{10}$$

where $\boldsymbol{r_0}$ is obtained from the one-point calibration procedure. h and k_0 are constants. These calibration parameters are obtained while a subject is looking around the PC screen *at will*. In order to determine h and k_0, the relationship between k_f and $|\boldsymbol{r}'|$ is plotted. The formula of the relationship is obtained by curve fitting. In this *nonlinear calibration*, $\boldsymbol{r_0}$ is used for both calibration and gaze detection, as seen in equation (10).

3 Experiments

3.1 Comparison of the Precision of Gaze Detection among the Three Calibration Methods: *one-calibration*, *two-calibration* and *nonlinear calibration*

Ten university students without eyeglasses served as the subjects of this experiment. The distance between the subject's face and the PC screen was approximately 80 cm. In the calibration procedure, the subject fixated on two calibration targets on center and top of the PC screen. The subject was asked to first look at the center calibration target for approximately two seconds, and to then look at the top calibration target. The obtained data for the top target was used only for the *two-point calibration* method [3] but not used for *one-point calibration* method. For *nonlinear calibration*, moreover, the subject was asked to gaze the 25 (five by five) calibration targets arranged on whole the PC screen in order to gaze whole the screen for the subject. The coordinates of the 25 targets were not used for calibration. After the calibration procedure, the subject fixated one by one on the 25 targets for approximately one second in order to compare the precision of gaze detection among the three calibration methods.

Fig. 6 shows the samples of the relationships between k_f and $|\boldsymbol{r}'|$, which are obtained from the experiment. In our system, the line of sight is obtained from the right and left eyes, respectively. Fig. 7 compares the average precision of the right and left gaze detection among the three calibration methods for each subject. The averages and SDs for *one-calibration* method were a gaze error in visual angle of 0.71 ± 0.41 degrees, whereas those of *two-point calibration* or *nonlinear calibration* were 0.68 ± 0.44 degrees and 0.59 ± 0.30 degrees, respectively.

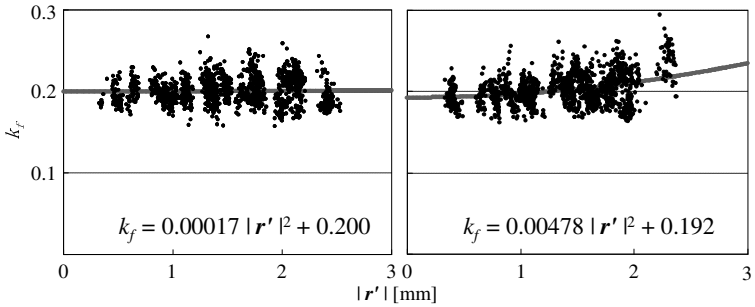


Fig. 6. Relationships between k_f and $|\boldsymbol{r}'|$

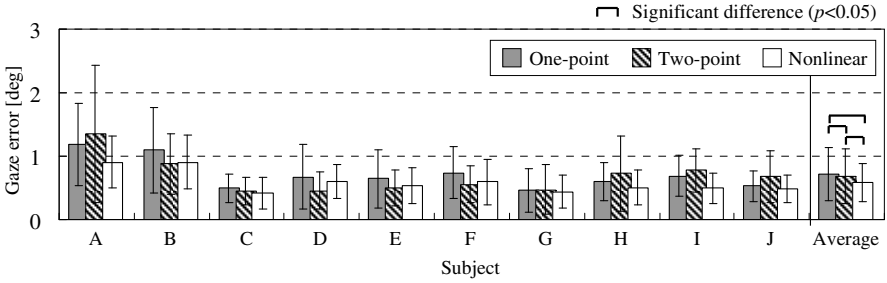


Fig. 7. Comparison of error of the gaze detection of both eyes among the three calibration methods

3.2 Gaze Calibration during Web Browsing

In experiment, for *one-point calibration*, the circle target was used. First the target was large. Suddenly, it shrank. This target attracted the attention of the subject, and made it easy to fixate on the target accurately. After the one-point calibration procedure, the subject was asked to look around the PC screen. During this nonlinear calibration, Google Maps was freely browsed on internet using a PC mouse. After this procedure, the error of gaze points when the subject fixated on the 25 targets was evaluated.

Fig. 8 shows the comparison in the detected gaze point distributions between the one-point and nonlinear calibration methods. You can see that there is a tendency that the nonlinear calibration method shows the errors smaller than the one-point calibration method, especially on the top of the screen. Fig. 9 (a) and (b) show the mean gaze errors distinguishing right and left eyes when each of two subjects fixated on the 25 targets. When the subjects fixated on ten targets presented on the upper part of the screen, it showed clearly difference in gaze error. We conducted another experiment for ten subjects, in which a subject was asked to the 25 targets to investigate the effect of the compensation in the nonlinear calibration method. The results resembled those of the above-mentioned experiment.

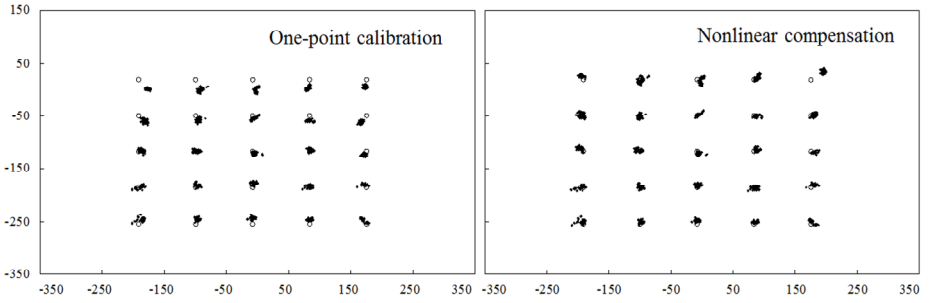


Fig. 8. Detected gaze point distributions between one-point and nonlinear calibration. Circles and dots indicate the target positions and the gaze points, respectively.

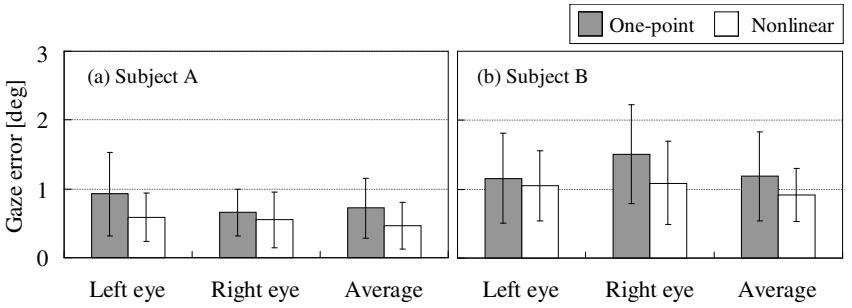


Fig. 9. Comparison in average gaze error between one-point and nonlinear calibration

3.3 Evaluation of Precision by Difference of Head Positions

The subjects were three university students. In the calibration procedure, the subjects were asked to fixate on the same calibration targets as in experiment 2 at approximately 80 cm from the PC screen. After the calibration procedure, the subjects fixated on the same 25 targets as in experiment 2 for the following nine head positions: approximately 70, 75, 80, 85, and 90 cm from the PC screen, and 5 cm to the left, 5 cm to the right, 5 cm to the top and 5 cm to the bottom at 80 cm. The subjects' heads were positioned using a chinrest stand. The calibration values that were obtained at the head position of 80 cm were commonly used for gaze detection at all head positions.

Fig. 10 shows the gaze errors when a subject changed the head position approximately 70, 75, 80, 85, and 90 cm from the PC screen, 5 cm to the left and 5 cm to the right at 80 cm, and 5 cm to the top and 5 cm to the bottom at 80 cm. The nonlinear calibration procedure was conducted at 80 cm. Except for both 5 cm to the top and 5 cm to the bottom, the average gaze error was 0.58±0.33 degrees in the new system while 0.92±0.40 [deg] in the previous system. The new system is improved by 35% in average compared to the previous system. The average gaze error of all head positions was 0.59±0.33 degrees in the new system. The new system uses the digital camera (640 by 480 pixel, 60 frame/sec, non-interlaced scanning) while the previous

used the NTSC analogue camera (640 by 240 field/sec, interlaced scanning). Accordingly, the difference of the cameras may have influenced the precision of gaze detection. However, this result implies that the new calibration method does not reduce the precision of gaze detection.

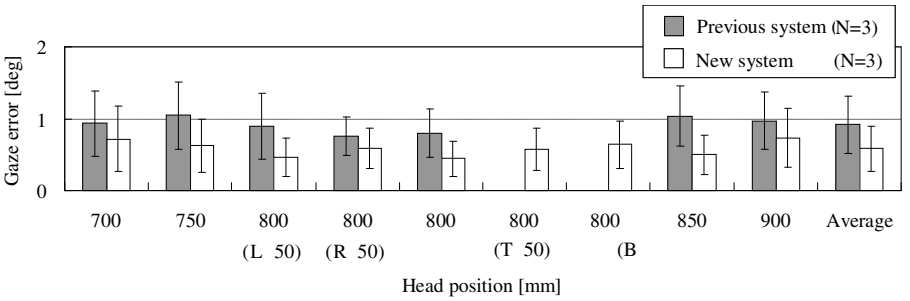


Fig. 10. Comparison in gaze error between our previous and new systems when subject displaced the head position

4 Conclusion

In conclusion, a subject must fixate on one specific target accurately even in the new calibration methods. However, the method improved the nonlinear relationship between θ and $|\boldsymbol{r}|$ without fixating on any specific position. Looking around in the PC screen is not burden for the subject. Also the new calibration method is very simple. Accordingly, the method is very useful in the experiment, in which the subject is an infant, who is difficult to fixate on a number of points on the screen accurately.

References

1. Model, D., Eizenman, M.: An Automatic Personal Calibration Procedure for Advanced Gaze Estimation Systems. IEEE Transactions on Biomedical and Engineering 57(5), 1031–1039 (2010)
2. Villanueva, A., Cabeza, R.: A Novel Gaze Estimation System With One Calibration Point. IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics 38(4), 1123–1138 (2008)
3. Ebisawa, Y., Abo, K., Fukumoto, K.: Head-Free, Remote Eye-Gaze Detection System with Easy Calibration Using Stereo-Calibrated Two Video Cameras. In: Stephanidis, C. (ed.) Posters, Part II, HCII 2011. CCIS, vol. 174, pp. 151–155. Springer, Heidelberg (2011)
4. Ebisawa, Y., Nakashima, A.: Increasing Precision of Pupil Position Detection Using the Corneal Reflection. The Institute of Image Information and Television Engineers 62(7), 1122–1126 (2008)