

A Widely Applicable Real-Time Mono/Binocular Eye Tracking System Using a High Frame-Rate Digital Camera

Keiji Matsuda¹(✉), Takeshi Nagami¹, Yasuko Sugase¹, Aya Takemura¹,
and Kenji Kawano²

¹ National Institute of Advanced Industrial Science and Technology (AIST),
Tsukuba, Japan

k.matsuda@aist.go.jp

² Kyoto University, Kyoto, Japan

Abstract. We have developed a new mono/binocular eye tracking system by using an IEEE1394b or USB-3.0 digital camera that provides high sensitivity, high resolution, high frame-rate and no rolling shutter distortion. Our goal is to provide a system that is friendly to researchers who conduct experiments. The system is non-invasive and inexpensive and can be used for mice, marmosets, monkeys, and humans. It has adopted infrared light to illuminate an eye (eyes). The reflected image of the infrared light on the cornea and the black image of the pupil are captured by the camera. The center of the pupil and the center of the corneal reflection are calculated and tracked over time. The movement of the head is compensated by the reflection. Since the high resolution camera has a 2048 horizontal pixels resolution, we can capture the images of both eyes simultaneously and calculate the parameters of the two eyes at each frame. The gaze position data can be read out on-line via computer network and/or DAC (digital analog converter). The adoption of the Windows 10 as the operation system makes this eye tracking system user-friendly. Because of the high frame-rate of the digital camera, the sampling rate of the system can be as high as 700Hz and the latency less than 4ms.

Keywords: Eye tracking · Image processing · Oculomotor system · High frame-rate digital camera

1 Introduction

The eye tracking system has been used for various purposes: in psychological studies, understanding visual information processing, analyzing consumers' interests, etc [1,2]. Recently, the eye movement information has also attracted a great deal of attention, because of its applicability in the Human-Computer-Interaction. However, as for the eye tracking system on the market, the hardware and software are incorporated as one manufactured product. As a result,

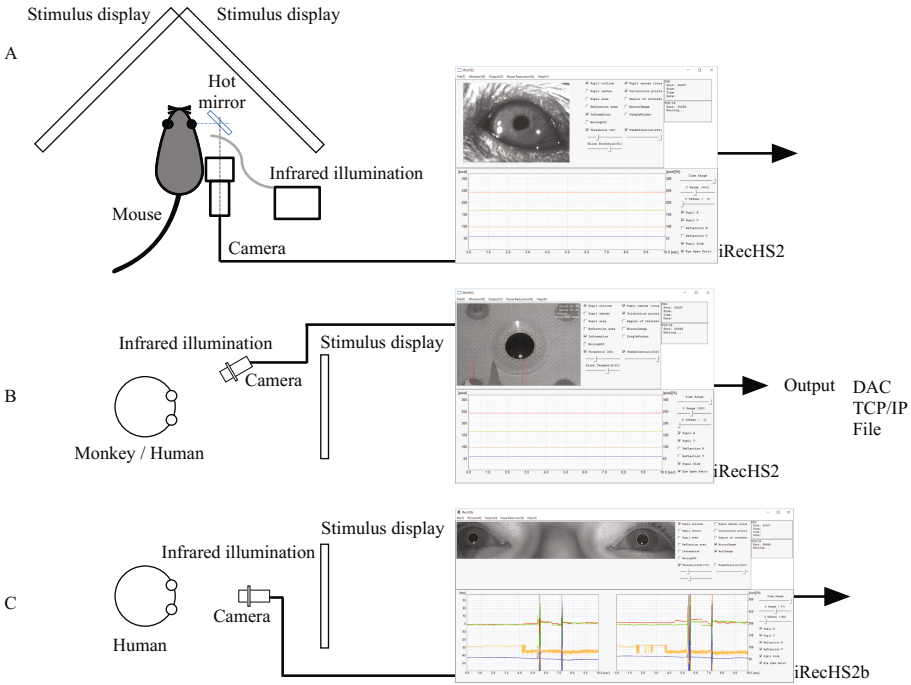


Fig. 1. Eye tracking system outline. In the monocular system (A) (B), the camera can be located at any position within the range where the pupil can be captured. In the binocular eye tracking system (C), the camera should be arranged at the same distance from each eye to be in focus.

the users (researchers) have to select or build up their experimental protocols to fit the eye tracking system that they own, which could limit their experiments. We report here a new user (experimenters) friendly, non-invasive and inexpensive eye tracking system, whose hardware can be selected to suit various experimental protocols. As shown in Fig. 1, there are two sets of software, iRecHS2 for monocular and iRecHS2b for binocular, both of which can be run on Windows 10. The basic requirement of the system’s hardware is a digital camera (FLIR Integrated Imaging Solutions, Inc.), which can receive infrared ray and has a global shutter. Other required hardware (infrared light-source, lens, PC, etc.) can be selected according to its application and the system can be used for mice, marmoset, monkeys, and humans, thus, expanding the applicability of the eye movement information over a variety of fields.

2 General Methods

Various types of eye tracking technologies are currently available [1,2]. In the present study, we selected a digital video-based combined pupil and corneal

reflection method by using a 3-D (3-Dimension) eyeball model, because it allows for passive calibration on untrained subjects (animals). By real-time image processing, the gaze angle, pupil size, blink (eye openness) and gaze-target position in 3-D space (binocular system) are estimated and the data can be read out on-line via TCP/IP and/or by a DAC (digital analog converter), and also be stored for offline analysis.

2.1 Apparatus

The system consists of the following devices. (1) infrared illumination, (2) visible light cut filter, (3) lens, (4) high frame-rate and global shutter digital camera, (5) Windows 10 personal computer, (6) DAC for output data and DIO (digital I/O) for synchronizing with other devices, (7) cables, camera support member, etc. The device configuration of (1)–(5) can be altered depending on the measurement subject and/or distance between the camera and subject. Details are summarized in Table 1.

Table 1. Equipments’ list.

	Mice	Monkey, human
Illumination	LA-100IR halogen lamp	IR LED array
Filter	62S PRO1D R-27 kenko	R-72 M30.5X0.5 Edmund optics
Lens	AIAF MicroNikkor 60 mm f/2.8D Nikon C mount Adapter kenko extender 2x	M7528-MP2 75 mm M3514-MP 35 mm CBC
Camera	GS3-U3-32S4M-C	
	GS3-U3-41C6NIR-C	
	FL3-U3-13Y3M-C (monocular only, not recommend)	
	GRAS-03K2M-C (monocular only, not recommend)	
	FLIR Integrated Imaging Solutions, Inc.	
PC	Windows 10 64 bit	
	Memory 16 GByte	
	Network interface	
	PCI Express x1 for DAC and DIO	
DAC with DIO	AO-1604L-LPE (4 ch DAC, 4 ch DIO for monocular)	
	AO-1608L-LPE (8 ch DAC, 4 ch DIO for binocular)	
	Cable PCB50PS-1.5P	
	Crimp relay terminal block EPD-50A	
	CONTEC	

2.2 3-D Eyeball Model for Gaze Detection

In order to obtain the gaze vector towards the camera, an eyeball model (Fig. 2) was adopted. In this model, the extension of the normal vector (C_x, C_y, C_z) of the pupil plane, passing through the center of the pupil, passes through the corneal curvature center and the eyeball rotation center. The vector (C_x, C_y, C_z) also represents the gaze vector in camera coordinates. We approximate the outline of the pupil in the captured image as an ellipse. We define the pupil center as (X_p, Y_p) , the length of the minor axis as S , the major axis as L , and the slope of the minor axis as a . Since the cornea is also a part of a sphere, we refer to the center of the sphere as the cornea curvature center (X_c, Y_c) . If the distance between the light/camera and the eyes is sufficiently longer than the head movement, the relationship between the center of the reflection and the cornea curvature center is constant. Thus, when the reflection point is (X_r, Y_r) , the cornea curvature center (X_c, Y_c) can be described as $(X_c, Y_c) = (X_r - Offset_x, Y_r - Offset_y)$. We defined the eyeball rotation center in a captured image as (X_o, Y_o) , length between the pupil center and eyeball rotation center in 3-D space as R and length between the pupil center and the corneal curvature center in 3-D space as R_{pc} . Using these parameters, we can express the gaze vector in two ways.

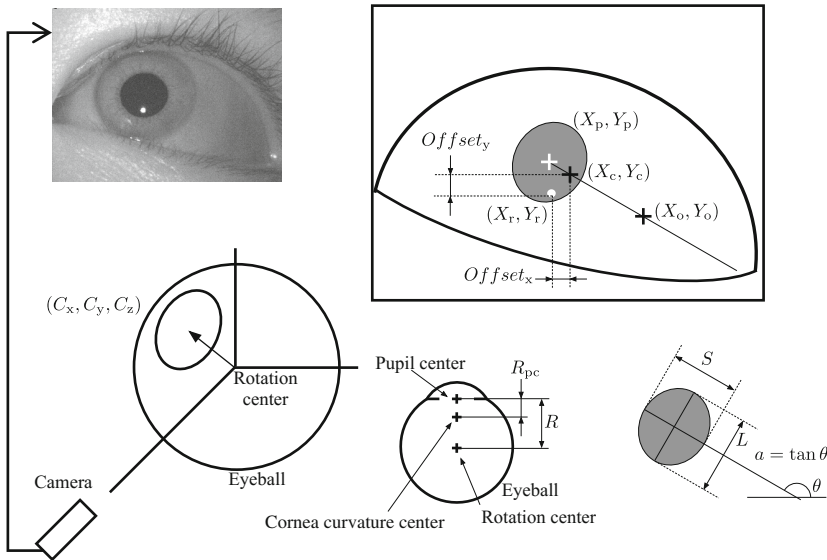


Fig. 2. 3-D eyeball model for gaze detection.

Head-fixed-method. Assess the gaze vector from the vector connecting the pupil center and eyeball rotation center (Eq. (1)). The gaze vector (C_x, C_y, C_z) can be calculated by using Eq. (1) after measuring the pupil center (X_p, Y_p) in the captured image of the eye. However, when the head moves, the eyeball rotation

center (X_o, Y_o) has to be calculated again. Thus, this method is applicable for the animal whose head is firmly fixed.

$$(C_x, C_y, C_z) = \left(\frac{X_p - X_o}{R}, \frac{Y_p - Y_o}{R}, \frac{\sqrt{R^2 - (X_p - X_o)^2 - (Y_p - Y_o)^2}}{R} \right) \quad (1)$$

Head-free-method. Assess the gaze vector from the vector connecting the pupil center and corneal curvature center (Eq. (2)). The gaze vector (C_x, C_y, C_z) can be calculated by using Eq. (2) after measuring the pupil center (X_p, Y_p) , and the gravity-center of the reflection in the captured image of the eye (X_r, Y_r) . By this method, the gaze vector can be calculated when the head moves in the picture frame. However, when the distance between the eye and camera changes, the image of the eye would be out of focus. Thus, using a chin-rest to keep the distance between them is preferable. Since the measurement can be taken only when the reflection light is on the cornea, the range of the measurable eye position is narrower than the head-fixed-method. The noise of the measurement might increase because the images of the pupil and reflection are used.

$$\begin{cases} X_c = X_r - Offset_x \\ Y_c = Y_r - Offset_y \\ (C_x, C_y, C_z) = \left(\frac{X_p - X_c}{R_{pc}}, \frac{Y_p - Y_c}{R_{pc}}, \frac{\sqrt{R_{pc}^2 - (X_p - X_c)^2 - (Y_p - Y_c)^2}}{R_{pc}} \right) \end{cases} \quad (2)$$

2.3 Pupil Detecting Method

In this system, an infrared illuminates the eye of the subject. By using a high frame-rate digital camera, the image of the subject's pupil is captured and approximated by an ellipse. To detect the pupil within limited time, we adopted the following procedure:

1. Capture an image from camera.
2. Create a reduced image of the captured image.
3. Create a binary image of the reduced image.
4. Label and measure the objects in the binary image. Detect the largest object as pupil.
5. Approximate the reduced image of the pupil by an ellipse.
6. Fit the ellipse on the original picture and detect the edge of the pupil.
7. Remove reflection of irrelevant objects on the pupil edge (Fig. 3).
8. Approximate the outline of the pupil to an ellipse by least square method.
9. Compare between the vertical height of the ellipse and the distance between the top and bottom of the pupil image to detect eye openness.
10. Create a binary image of the reduced image by using the threshold of the reflection.

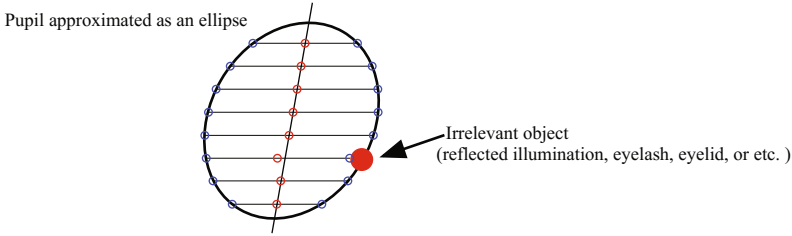


Fig. 3. Removal method for irrelevant reflected objects. Blue circles indicate detected pupil edges. Red circles indicate center of horizontal line segments. The points out of the straight line are omitted for the further calculation. (Color figure online)

11. The reflected image of the light source nearest to the pupil center is used to calculate the gravity-center.

The process between (2)–(5) can be skipped because of the small shift of the pupil during the high frame-rate sampling.

2.4 Passive Calibration

By applying the value of the captured images of the freely moving eye, the pupil rotation center (X_o, Y_o) and pupil rotation radius R are calculated using the head-fixed-method, and the offset between reflection and the corner curvature center $(Offset_x, Offset_y)$, and the length between the pupil center and corner curvature center R_{pc} are calculated using the head-free-method. When the slope of minor axis of the ellipse is defined as a , (X_o, Y_o) is expressed by Eq. (3). As shown in Fig. 4, the intersections of the minor axes of the pupil-ellipses in multiple picture-frames are located on (X_o, Y_o) . $Offset_x$ and $Offset_y$ are calculated by Eq. (4) using data taken from two or more picture-frames (least square method). R is calculated by Eq. (5) with (X_o, Y_o) and (X_{pi}, Y_{pi}) . R_{pc} is calculated by Eq. (6) with $Offset_x, Offset_y, (X_{pi}, Y_{pi}), (X_{ri}, Y_{ri})$. The calculation-process for R, R_{pc} from the major axis L and minor axis S of the pupil is shown in Fig. 5. By acquiring these parameters, the gaze vector directing to the camera can be calculated. This method is used for the subject who cannot perform the active calibration, for example mouse.

$$\begin{cases} a_i(X_o - X_{pi}) = Y_o - Y_{pi} \\ i = 0 \dots n \end{cases} \tag{3}$$

$$\begin{cases} a_i(X_{ri} - Offset_x - X_{pi}) = Y_{ri} - Offset_y - Y_{pi} \\ i = 0 \dots n \end{cases} \tag{4}$$

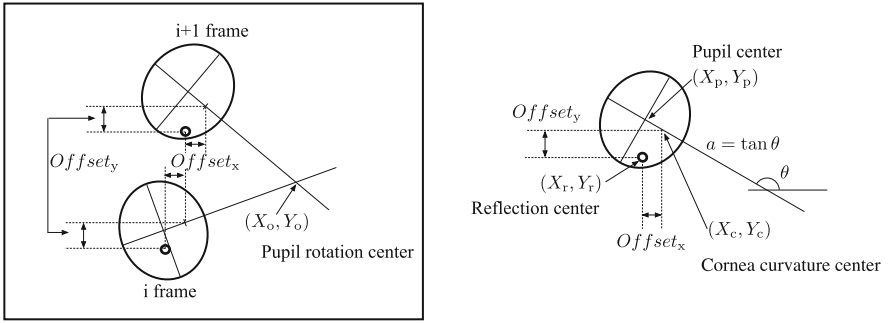


Fig. 4. A schematic representation of the method to calculate (X_o, Y_o) , $Offset_x$, $Offset_y$

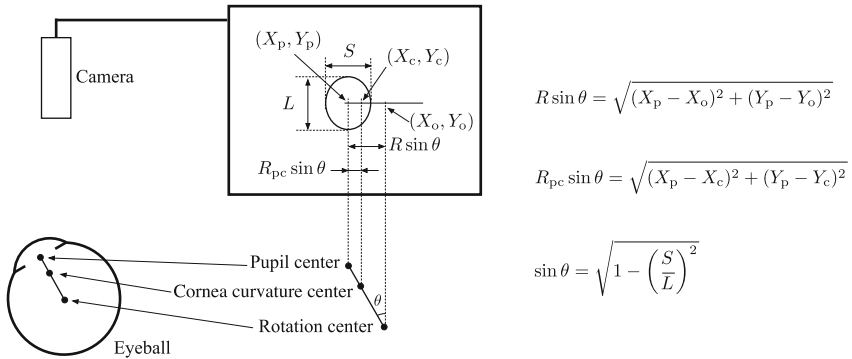


Fig. 5. A schematic representation of the method to calculate R , R_{pc} .

$$\begin{cases} R = \sqrt{\frac{(X_{pi} - X_o)^2 + (Y_{pi} - Y_o)^2}{1 - \left(\frac{S_i}{L_i}\right)^2}} \\ i = 0 \dots n \end{cases} \quad (5)$$

$$\begin{cases} R_{pc} = \sqrt{\frac{(X_{pi} - X_{ri} + Offset_x)^2 + (Y_{pi} - Y_{ri} + Offset_y)^2}{1 - \left(\frac{S_i}{L_i}\right)^2}} \\ i = 0 \dots n \end{cases} \quad (6)$$

2.5 Active Calibration, Transformation into Object Coordinates

We calculate a gaze axis vector (C_x, C_y, C_z) in camera coordinates from Eqs. (1), (2). and obtain a gaze axis vector (T_x, T_y, T_z) in object coordinates

from target positions. The subject is instructed to fixate small targets, located at 5 (more than 3) positions on the stimulus display to obtain $Offset_x, Offset_y, R_{pc}$, and a 3×3 matrix M for coordinates transformation (Eq. (7)). Then we can calculate a gaze vector (O_x, O_y, O_z) in object coordinates from Eq. (8).

$$M = \begin{pmatrix} \sum_{i=1}^n C_{xi}^2 & \sum_{i=1}^n C_{xi}C_{yi} & \sum_{i=1}^n C_{zi}C_{xi} \\ \sum_{i=1}^n C_{xi}C_{yi} & \sum_{i=1}^n C_{yi}^2 & \sum_{i=1}^n C_{yi}C_{zi} \\ \sum_{i=1}^n C_{zi}C_{xi} & \sum_{i=1}^n C_{yi}C_{zi} & \sum_{i=1}^n C_{zi}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^n C_{xi}T_{xi} & \sum_{i=1}^n C_{xi}T_{yi} & \sum_{i=1}^n C_{xi}T_{zi} \\ \sum_{i=1}^n C_{yi}T_{xi} & \sum_{i=1}^n C_{yi}T_{yi} & \sum_{i=1}^n C_{yi}T_{zi} \\ \sum_{i=1}^n C_{zi}T_{xi} & \sum_{i=1}^n C_{zi}T_{yi} & \sum_{i=1}^n C_{zi}T_{zi} \end{pmatrix} \tag{7}$$

$$\begin{pmatrix} O_x \\ O_y \\ O_z \end{pmatrix} = M \begin{pmatrix} C_x \\ C_y \\ C_z \end{pmatrix} \tag{8}$$

2.6 Case of Binocular Vision

We calculate gaze-target positions (x, y, z) in 3-D space from the vectors in object coordinates of two eyes (left (LO_x, LO_y, LO_z) , right (RO_x, RO_y, RO_z)) and inter-pupillary distance (IPD) (Eq. (9)).

$$\begin{cases} \frac{x + \frac{IPD}{2}}{LO_x} = \frac{y}{LO_y} = \frac{z}{LO_z} \\ \frac{x - \frac{IPD}{2}}{RO_x} = \frac{y}{RO_y} = \frac{z}{RO_z} \end{cases} \tag{9}$$

3 Evaluation of the System’s Accuracy

3.1 Evaluation of the System’s Accuracy of Gaze Angle

We evaluated the system’s accuracy by using an apparatus of synthetic eyes. (Fig. 6).

Method. We measured the accuracy of the system in representing monocular gaze angles. First, we calibrated the system while the synthetic eye gazing at 8 points angle of $(horizontal, vertical) = (-10, 10), (0, 10), (10, 10), (-10, 0), (0, 0), (10, 0), (-10, -10), (10, -10)$ degrees. Then, under the head fix method, $(X_o, Y_o), R$, and translation matrix M_{fixed} were calculated, and under the head-free-method, $Offset_x, Offset_y, R_{pc}$, and translation matrix M_{free} were calculated. We set the synthetic eye gazing at 25 points, horizontal angle -10 to $+10^\circ$ and vertical angle -10 to 10° at intervals of 5° , and measured the eye’s gaze angle for 1 s at each setting. The camera was GS3-U3-41C6NIR-C, and the captured image size was 320×240 pixels.

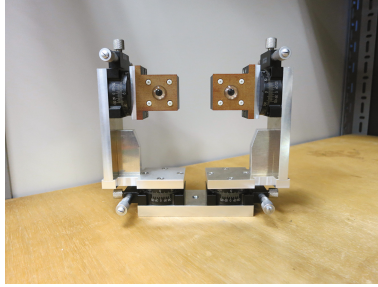


Fig. 6. A pair of synthetic eyes. The cornea was acrylic. The diameter of the pupil: 5 mm. The radius of cornea curvature: 7.7 mm. The radius of rotation of the pupil center: 9 mm. The distance between two eyes: 60 mm. The axis of rotation (both X, Y-axis): high-grade transmission type rotary stage (CHUO PRECISION INDUSTRIAL CO., LTD. RS-4012-R, vernier reading 10 min of arc).

Result. Figures 7, 8 show the distribution of the gaze angles in the head-fixed-method (Fig. 7) and head-free-method (Fig. 8). The gaze angles are plotted in camera coordinates (A) and in object coordinates (B), and superimposed in both coordinates (C). The distribution of the gaze positions in camera coordinates shrank a little from those in object coordinates (C). The means and standard deviations of the errors at 25 points (in object coordinates) are summarized in Table 2. Figure 9 shows the data at the center in a larger scale (A, B) and those after a median filter. The scatter of the errors by using the head-free-method was larger than that by the head-fixed-method.

Discussion. As shown in Figs. 7C and 8C, the distribution of the gaze positions in camera coordinates shrank a little from those in object coordinates. It is because of the shift of the pupil center due to the cornea-refraction. It is

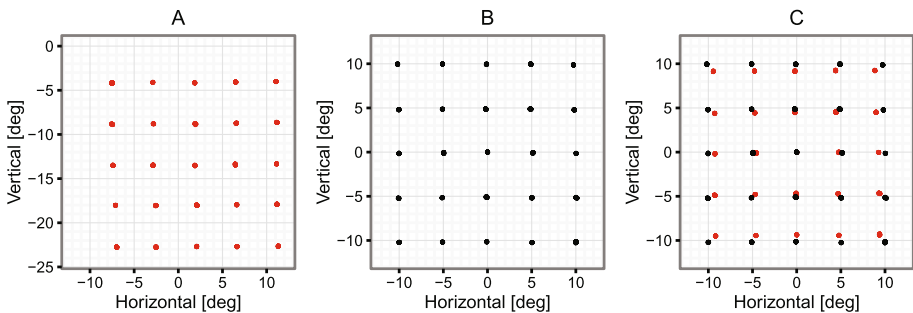


Fig. 7. Head-fix-method. A: After passive calibration, gaze angle in camera coordinates. B: After active calibration, gaze angle in object coordinates. C: The data in A and B are superimpose after parallel shift.

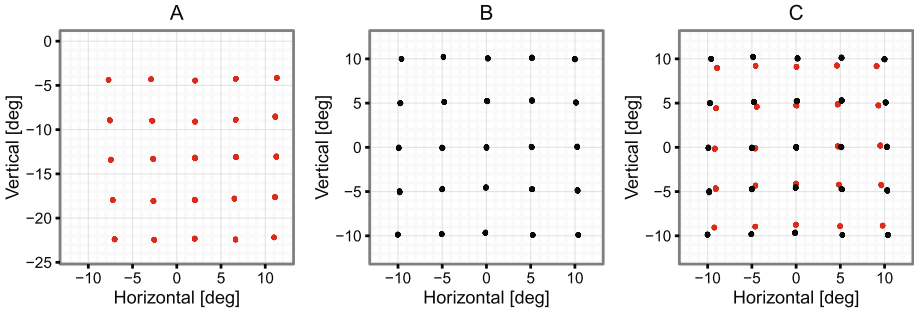


Fig. 8. Head-free-method. A: After passive calibration, gaze angle in camera coordinates. B: After active calibration, gaze angle in object coordinates. C: The data in A and B are superimpose after parallel shift.

Table 2. System’s accuracy of gaze angle.

	Head-fixed-method [mean(sd)]	Head-free-method [mean(sd)]
Horizontal	0.028 (0.098) degree	-0.042 (0.138) degree
Vertical	0.096 (0.071) degree	0.047 (0.139) degree

necessary to consider the shift when the subject (such as a mouse) cannot perform active calibration. On the other hand, there is no practical issue for the subject, who can perform the active calibration, because the transformation matrix created by the active calibration compensates the shift. The larger scattering of the errors in the head-free-method (Fig. 9A, B) is due to the measurement error of the gravity-center of the reflection in addition to the measurement error of the pupil center. When the head is fixed perfectly, the head-fixed-method provides higher-accuracy and wider-range measurement. Practically, the researchers can get stable measurements, by applying a median filter or moving average filter (Fig. 9C, D).

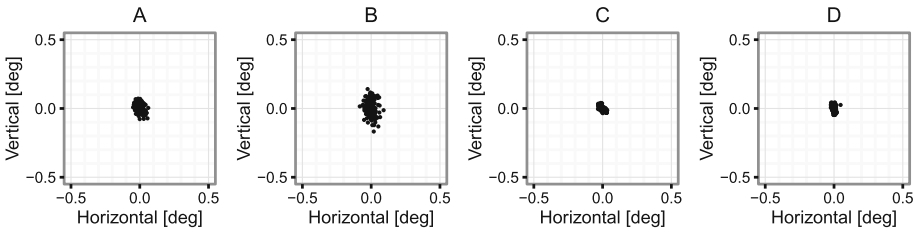


Fig. 9. Distribution of eye position data at the center of fixation point (0,0). The X- and Y-axis scale are expanded. A: Superimposed eye position data taken by the head-fixed-method. B: Taken by the head-free-method. C: By the head-fixed-method after applying a median filter. D: By the head-free-method after applying a median filter.

3.2 Evaluation of the System’s Accuracy of Gaze Position

Method. We estimated the accuracy of the system in representing binocular gaze angles by using a pair of synthetic eyes. First, we calibrated the system while the synthetic eyes gazing at 8 points horizontal angles of $(righteye, lefteye) = (1, -1), (1.5, -1.5), (2.0, -2.0), (2.5, -2.5), (3.0, -3.0), (3.5, -3.5), (4.0, -4.0), (5.0, -5.0), (6.0, -6.0), (8.0, -8.0), (10.0, 10.0)$ degrees, and vertical angle of 0° . The horizontal position at $(1, -1)$ indicates that the gaze is on the center of the visual field at distance of 1719 mm, and that at $(10, -10)$ indicates at distance of 170 mm. We measured the left and right eyes’ gaze angles and calculated the 3-D position of the intersection for 1 s at each setting.

Result. Figure 10 shows the distribution of eye position errors (A, B) and 3-D intersections (C) measured by head-free-method, by using a 10-points moving average filter.

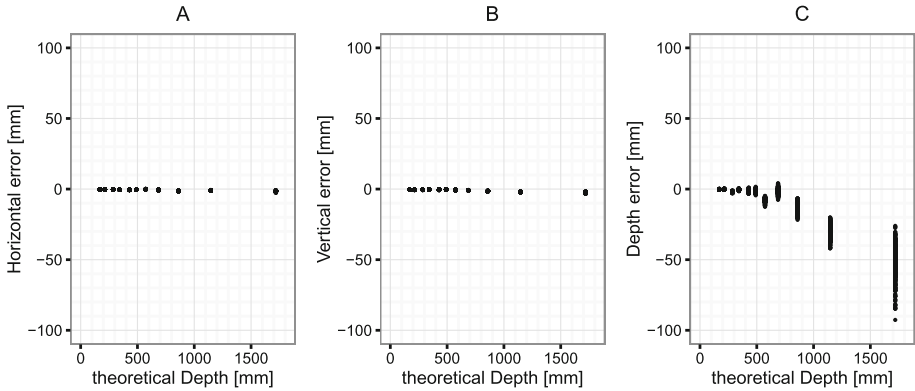


Fig. 10. Distribution of error of position (A, B) and 3-D intersection. A: horizontal B: vertical C: depth.

Discussion. As shown in Fig. 10, when the distance between the subject and the target is larger, the distribution of the measurement errors of the intersection scattered. It is important to notice this characteristic of the 3-D measurement.

4 Evaluation of the System’s Latency

We evaluated the system’s latency by using an apparatus of synthetic eye (Fig. 11).

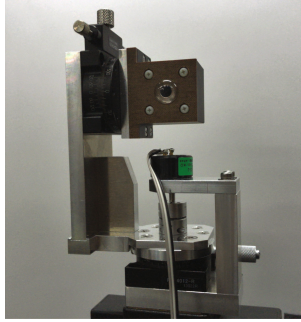


Fig. 11. A potentiometer (Midori Precision CP-2F(b), linearity $\pm 1\%$) was attached to the synthetic eye. The rotation angle can be measured by applying voltage to its edges.

4.1 Methods

By using the synthetic eye, the system was calibrated. The voltage output from the potentiometer and eye tracking system was sampled at 10 kHz. The synthetic eye was rotated by a hand and the output voltage were recorded and translated into rotation angle and horizontal eye movements (Fig. 12). As shown by the enlarged view in Fig. 12B, D, the time-delay of the system can be observed. The latency of the system is calculated by the following Eq. (10), by using the rotation angle output of potentiometer as $p(t)$ and the output of the eye tracking system as $h(t)$. The latency was obtained when the Error RMS (root mean square) was the minimum. We measured latencies with different sampling frequencies of the picture-frame.

$$Error\ RMS = \sqrt{\frac{1}{n} \sum_{i=0}^n \{p(t_i) - h(t_i + latency)\}^2} \quad (10)$$

4.2 Results

The data calculated by using Eq. (10) are shown in Fig. 13A at seven sampling frequencies of the camera-frame. Figure 13B shows the relationship between the minimum value and sampling frequency, suggesting that the system's latency depended on the sampling frequency ($= 1/\text{period}$). The approximated line in Fig. 13B indicates that the latency is represented as the sum of the sampling duration and processing time. The latency was less than 4 ms at 500 Hz sampling rate.

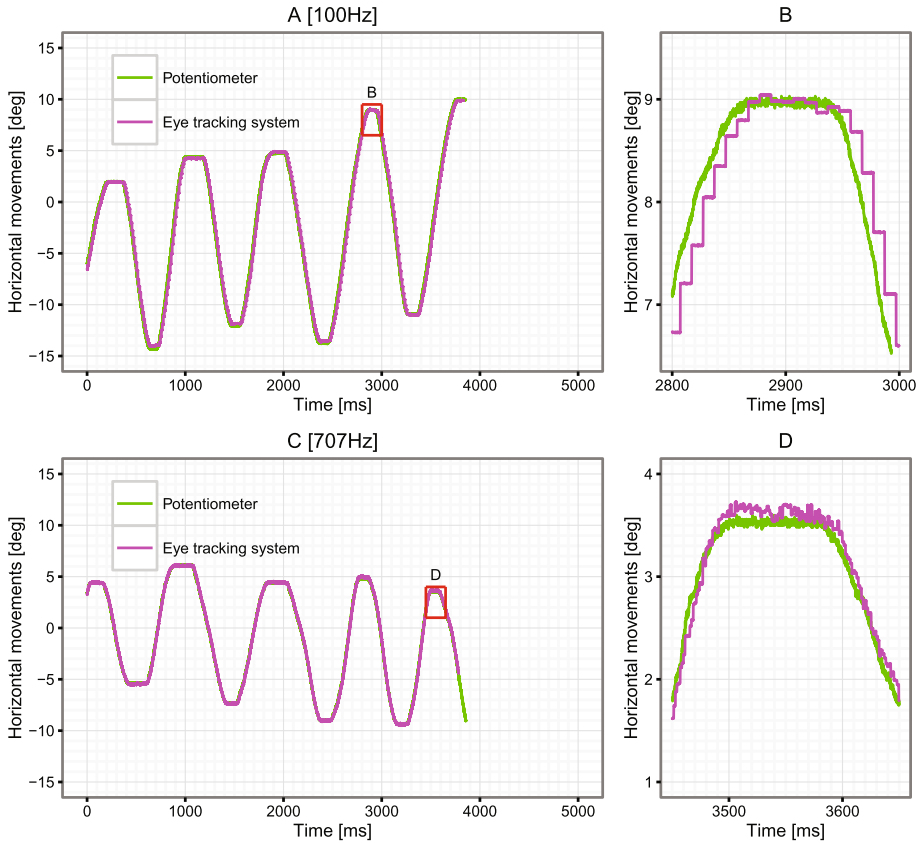


Fig. 12. An example record by using a GS3-U3-41C6NIR-C with two sampling frequencies of the picture-frame (A, B at 100 Hz, and C, D at 707 Hz). Green line indicate potentiometer rotation angle and magenta line the system's output (horizontal eye position). (Color figure online)

4.3 Discussion

The reliable measurement of the system's latency is one of the advantages of this eye tracking system. To get stable measurement, we have suggested application of a median filter or moving average filter. Since a 9-points median filter causes a 4.5-points time delay, the time-delay at the sampling frequency of 500 Hz is estimated as 13 ms. Thus, it is possible to calculate the gaze angle during the period of one frame and to display a new picture on the next-frame of a 60 Hz video display (real-time eye-position feedback experiment).

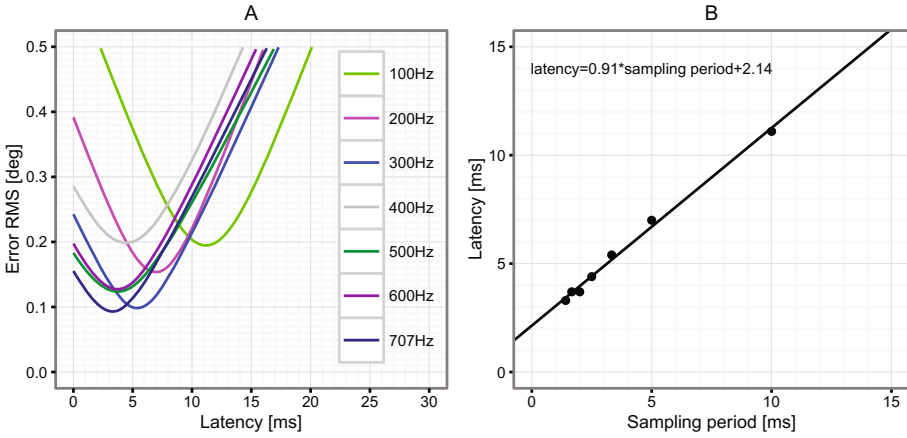


Fig. 13. A: The data calculated from by using Eq. (10). B: The relationship between the minimum value and sampling period. The sampling frequency are 707, 600, 500, 400, 300, 200, 100 Hz.

5 Measuring of Human Gaze Traces in 3-D Space

5.1 Methods

The subject was an adult male wearing glasses. The subject’s head was fixed by a chin rest. The system was calibrated by measuring eye positions during the subject fixed targets on the display. After the calibration, the subject transferred his gaze in 5 ways, between the center position (near) to one of the 5 positions (far) (Fig. 14). The sampling frequency was 500 Hz.

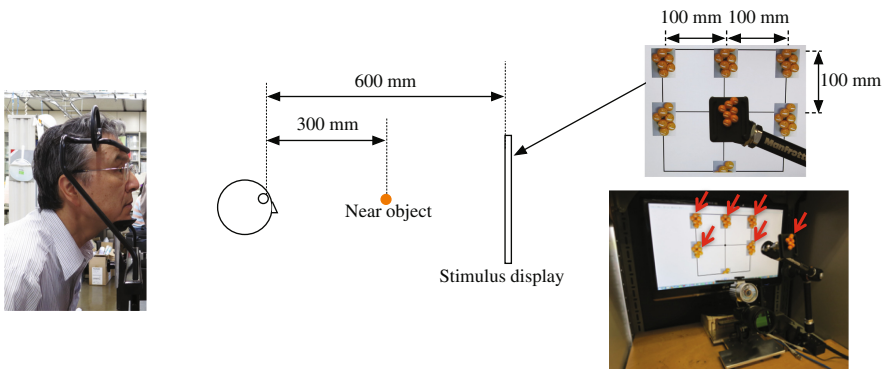


Fig. 14. Schematic representation of the human experiment and visual stimuli in 3-D space. Red arrows indicate one near (at 300 mm) and 5 far (at 600 mm) targets. (Color figure online)

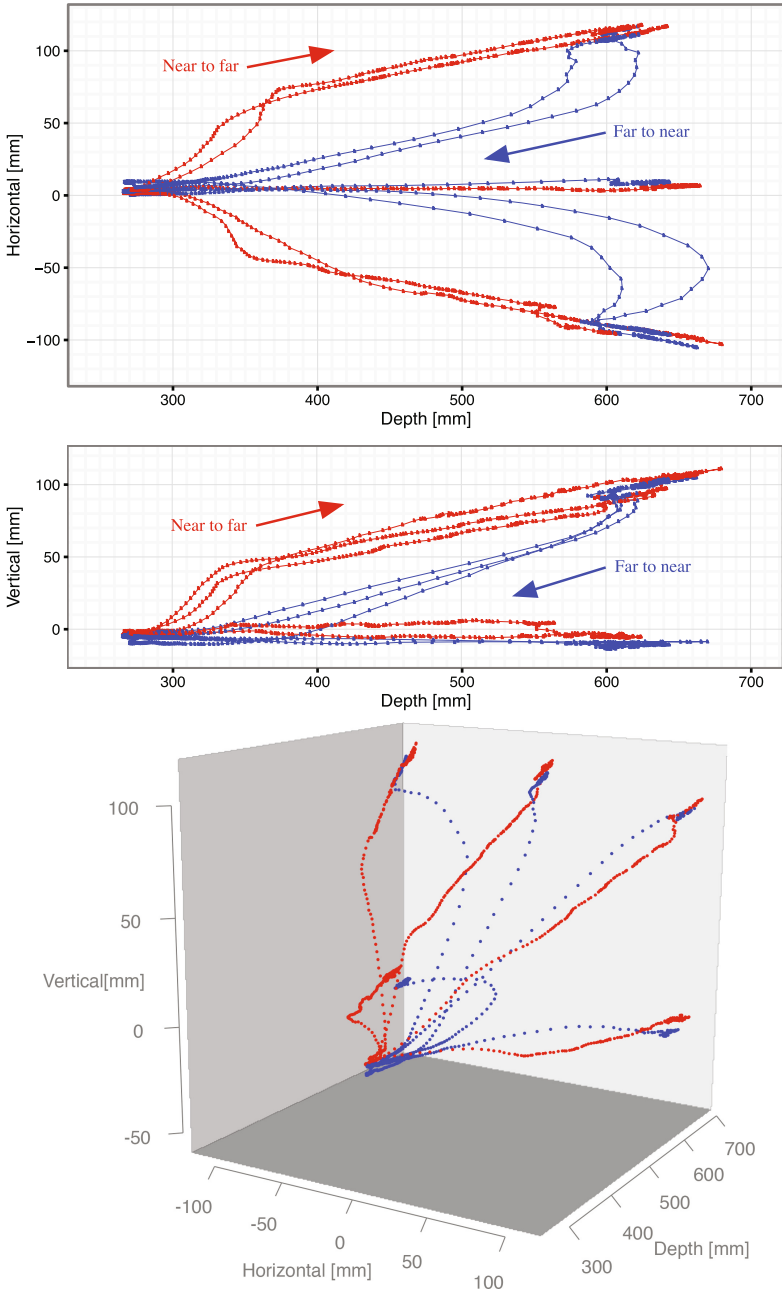


Fig. 15. The gaze transfer traces. (Color figure online)

5.2 Results

Figure 15 shows the gaze transfer traces after applying a 10-points (20 ms) moving-average-filter. It clearly shows that the eye movement traces from near to far (divergence, red trace in Fig. 15) and from far to near (convergence, blue trace) are different in 3-D space. In both cases, the vergence eye movements (convergence and divergence) started earlier than saccade and terminated after the saccades.

5.3 Discussion

The gaze transfers between the near to far positions consisted of combination of saccades and vergence eye movements. The subject transferred his gaze from the near to far (red dots) point and from the far to near (blue dots) point in different ways. In both cases, the vergence eye movements started earlier than saccades. The results are consisted with classical reports by Yarbus (1957) [3] and Enright (1984) [4].

6 Conclusion and General Discussion

We developed a new widely applicable eye tracking system. By using synthetic eyes, we found that the system can measure eye movements with better than 0.2° accuracy and latency was less than 4 ms at 500 Hz. By using this system, eye movements can be measured even when the subject (such as a mouse) cannot perform the active calibration. Since this system takes the images of two eyes by one camera, movements of two eyes can be measured simultaneously. With this system, we succeeded in characterizing human vergence/saccade eye movements when ocular fixation shifted between two targets placed at different distances in 3-D space.

These programs can be access from <https://staff.aist.go.jp/k.matsuda/iRecHS2/>.

Acknowledgement. This paper is based on results obtained from a project commissioned by KAKENHI (24650105, 16H03297) and the New Energy and Industrial Technology Development Organization (NEDO).

References

1. Duchowski, A.: Eye Tracking Methodology: Theory and Practice. Springer, London (2007)
2. Holmqvist, K., Nystrom, M., Andersson, R., Dewhurst, R., Jarodzka, H., Tracking, E.: A Comprehensive Guide to Methods and Measures. Oxford University Press, USA (2011)
3. Yarbus, A.L.: Eye movements during changes of the stationary points of fixation. *Biophysics* **2**, 679–683 (1957)
4. Enright, J.T.: Changes in vergence mediated by saccades. *J. Physiol.* **350**, 9–31 (1984)