2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

Toward Micro Eye Movement Detection in Practice: Stand-alone Eye Tracker with High Resolution and Wide Measurement Range

Keiko Yokoyama[1], Tomohiro Sueishi[2], Michiaki Inoue[3],
Shoji Yachida[1], Toshinori Hosoi[1] and Masatoshi Ishikawa[2]

Abstract—
Detecting the micromovements of eyes that reflect a person's inner state can be an essential step in many applications, but most eye trackers need to fix the subject's head using a chin rest to obtain sufficient data quality. We propose a stand-alone eye tracker that utilizes two 500 fps cameras, a pair of rotating mirrors for gaze control, a liquid lens for focus control and an intensity-controllable light source, and describe how the proposed system works in real-time. The experimental results show that our system covers more than twice as wide a measurement range in the depth direction as the conventional eye tracker while achieving sufficient data quality to analyze microsaccades with an amplitude of down to 0.2 deg. We also examine the practical use of the proposed system for microsaccade detection.

I. INTRODUCTION

As the old saying goes: "Eyes are the windows to the soul." The state of the eyes embodies not only a person's action intentions but also inner states of mind. For example, eye gaze information can be utilized for controlling or improving the functionality of robots [1] [2]. Moreover, distinctive eye movements such as saccades (rapid eye movements that occur when shifting gaze to another object) and microsaccades (small saccades that occur during attempted fixation) can identify a wide variety of internal conditions, including concentration [3], depression [4], and even neurological / cognitive disorders [5][6]. There will be many possible applications by combining them, however, measuring small eye movements like microsaccades in real-world environments is difficult due to device limitations.

Devices for measuring eye movements are called eye trackers. Currently, most eye trackers are based on a non-invasive method utilizing one or more cameras, and are broadly divided into head-mounted types and stand-alone types. For measurement of small eye movements, the stand-alone type is suitable because of the required spatio-temporal resolution. However, most stand-alone-type eye trackers assume that neither the device and the head moves during measurements, especially in the depth direction, with the head fixed on a chin rest. This would not only be stressful for the subject but also would limit the use of the device.

[1] Keiko Yokoyama, Shoji Yachida and Toshinori Hosoi are with Visual Intelligence Research Laboratories, NEC, 1753 Shimonumabe, Nakahara-ku, Kawasaki-shi, Kanagawa, Japan. k.yokoyama@nec.com

[2] Tomohiro Sueishi and Masatoshi Ishikawa are with Research Institute for Science, Tokyo Univ. of Science, 6-3-1 Niijuku, Katsushika-ku, Tokyo, Japan. sueishi@ishikawa-vision.org

[3] Michiaki Inoue is with Digital Solutions Division, NEC Solution Innovators, Ltd., 2-6-1 Kitamigata, Kawasaki-shi, i, Takatsu-ku, Tokyo, Japan. michiaki-inoue@nec.com

There are two possible ways to make the stand-alone eye tracker robust against head movements in the depth direction: to increase the spatial resolution of the eye images and to deepen the depth of field to achieve better focus. However, high resolution over a wide area leads to an enormous processing load, and closing down the aperture for a greater depth of field will darken the image.

This paper proposes a stand-alone eye tracker that enables microsaccade measurements while allowing dynamic head movement against the device. The proposed system consists of two 500 fps cameras and has following features.

- A mechanism to control the camera gaze, focal length, and light intensity with low latency.
- Light-weight algorithms to update the control values and estimate the eye angle in real-time.

We evaluate its performance in terms of the measurement range in the depth direction and the robustness against head movement. We also discuss the most effective way to detect microsaccadic movements in the measured data.

II. BRIEF REVIEW OF EYE TRACKING

A. Eye Tracking Principles

Eye tracking systems can be broadly classified into three categories: electro-oculography (EOG), scleral search coil method, and video-oculography (VOG). EOG measures the myoelectric potentials generated by eye movements with a myoelectrometer attached to the face. This method is quite simple but has low spatial and temporal resolution. The scleral search coil method is highly accurate thanks to recording small electric currents induced by the magnetic field in a coil embedded in a donut-shaped plastic ring that is placed over the eye. However, it causes discomfort for patients due to its invasiveness. VOG is the most commonly utilized method these days, as it achieves both non-invasiveness and high-accuracy thanks to using stereo cameras.

B. Eye Angle Measurement Method

Most VOG-based eye trackers harness the pupil and corneal reflection (CR) principle [7], where eye rotations are inferred from pupil-CR vectors with more than one light source in known locations in combination with a 3D model of the eyeball. Usually, the eyeball is represented as a model consisting of two spherical surfaces with two different radii of curvature (Fig. 1). The obtained images are divided into two types depending on how the illumination is installed: one aligns the optical axes of the camera and illumination and extracts the pupil as a bright area (bright pupil method), and

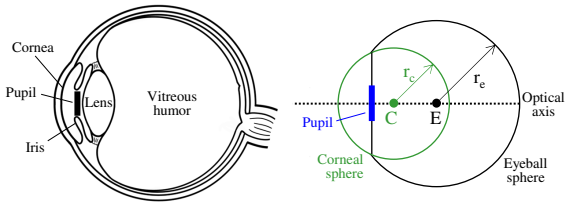

Fig. 1. Eye anatomy (left) and its simple model (right).

the other positions the light source away from the camera optical axis and extracts the pupil as a dark area (dark pupil method). The dual-Purkinje-image (DPI)-based approach [8], which utilizes the lens reflection instead of the pupil, has also been proposed to improve the accuracy over the pupil-CR method. However, the lens reflection has an extremely low luminance compared to the pupil and cornea, making it impractical.

C. VOG-based Eye Trackers

There are roughly two types of VOG-based eye trackers: head-mounted and stand-alone. Head-mounted eye trackers are built into a headband or glasses that are worn by the subject, allowing for a high degree of postural freedom. However, they usually do not have enough temporal or spatial resolution to measure fast and small eye movements (e.g., microsaccades). As for stand-alone eye trackers, many of them utilize a position-calibrated flat screen as the stimulus area, have higher precision and are more suitable for microsaccade research. In general, however, there is a trade-off between data quality and robustness to head movement. For example, EyeLink 1000 Plus [9] is one of the most commonly used eye trackers for microsaccade research due to its high sampling rate (up to 2,000 Hz) and precision, but the allowed head movement is only 2 cm in depth. Another state-of-the-art eye tracker, Tobii Pro Spectrum 1200 [10], is more flexible and tolerant to head movement (up to 20–30 cm in depth), but it is still not sufficient when considering natural head movement in the sitting position. According to the statistical data of the dimensions of men and women sitting [11], the eye level from the seat height is around 750 mm. Then, if a sitting person leans back and forth ± 12 deg, the eye moves more than 300 mm in the depth direction.

Various eye trackers based on active vision have been proposed to tackle this issue. For example, Yoo and Chung [12] proposed using two cameras, a wide-angle camera and a telephoto camera, mounted on a pan-tilt unit to keep the subject's face in the field of view. In another example, Kim et al. [13] utilized a pair of rotating mirrors instead of a pan-tilt unit to control the camera gaze. However, these systems run at only 10–15 Hz, which is too weak for microsaccade measurements.

In light of this background, we propose an eye tracker with sufficient temporal and spatial resolution for microsaccade measurements while setting subjects more free from head position restraint. Specifically, given that the microsaccade duration is 6–25 ms [14] and the average magnitude is 0.5 ± 0.3 deg [15], we aim to achieve a system capable of measuring eye displacements equivalent to 0.2 deg at a temporal resolution of 500 fps.

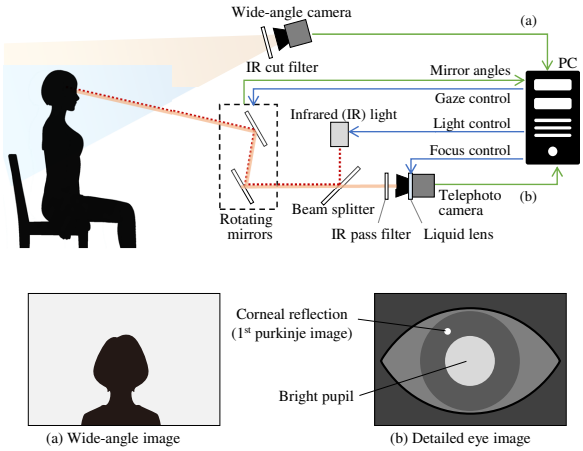

Fig. 2. Overview of proposed eye tracking system.

III. ROBUST EYE TRACKER FOR DYNAMIC HEAD MOVEMENT

A. System Model

Fig. 2 shows an overview of the proposed eye tracking system. In this system, a wide-angle camera captures eye pairs in the face area while a telephoto camera, with a pair of rotating mirrors in front, tracks one of the eye pupils that looks brighter than the other areas by retro-reflecting infrared light coaxial to the camera. Simultaneously, the depth position of the target eye is calculated using a triangulation algorithm to adjust the focus of a liquid lens attached to the telephoto camera, and the light intensity is adjusted to maintain constant pupil brightness in the image.

The key concept of our system is to control gaze, focus, and light intensity in real-time with low latency so that the telephoto camera continuously obtains focused images with the eye pupil always at the image center. This enables stable measurement over a wide range—not only in the vertical and horizontal directions but also for head movement in the depth direction, which conventional stand-alone eye trackers have difficulty with.

B. Process Flow

Initially, the target eye position in the wide-angle camera image is manually instructed on the GUI to adjust mirror angles so that the eye moves within the angle of view of the telephoto camera. Then, the proposed system performs the process shown in Fig. 3 in real time, synchronized with the camera frame rate of 500 fps. The target eye is detected in the wide-angle camera image using a hierarchical scheme combined with face and eye detection [16], and the bright pupil and CR of the eye in the telephoto camera image are tracked to calculate pupil-CR vectors by using a self-windowing algorithm [17], a lightweight tracking method for high-speed cameras that limits the possible region where the target exists under the assumption of a high frame rate. To keep the eye within the field of view, the angles of the

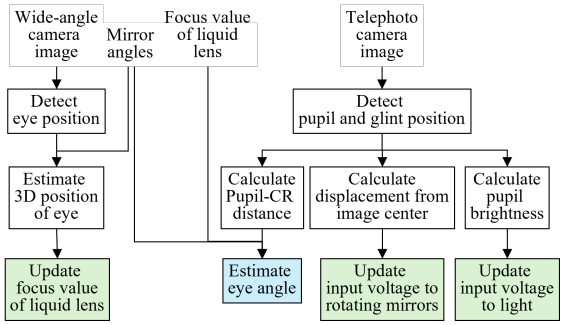

Fig. 3. Main process flow diagram of proposed system. Initial detection of the target eye is to be performed beforehand.

rotating mirrors (i.e. gaze angles of the telephoto camera) are controlled by proportional-integral-derivative (PID) control with the displacement of eye pupil center from the image center. The light intensity is also controlled by PID control to maintain a constant average luminance value within the pupil area. The focus control value for the liquid lens is updated in accordance with the depth position of the eye, which is derived from the eye position in the wide-angle camera and the mirror angles. Then, the eye angle is estimated from the pupil-CR vector by taking into account the system-specific factor (i.e., magnification rate change) and head translation.

C. Prototype

Fig. 4 shows a prototype of the proposed system, which features the folliwing components.

- A camera (Basler, ace acA800-510um, $720 \times 500[\mathrm{px}^2]$, 500 fps) with a wide-angle lens (Edmond Optics, 6mm C Series Fixed Focal Lens)
- A camera (FLIR, Blackfly S, $720 \times 540[\mathrm{px}^2]$, 500 fps) with a telephoto lens (Nikon, AI AF Zoom-Nikkor 80–200 mm, focal length fixed at 105mm) and a liquid lens (Optotune, EL-16-40-TC-VIS-20D-1-C)
- A pair of rotating mirrors (Thorlabs, VantagePro, scanning angle $\pm 22.5\,\mathrm{deg}$)
- A collimated infrared light (Leimac, IBF-LXS30AIR-850, wave-length 850 nm), whose optical axis is set as coincident with that of the telephoto camera
- A computer (GALLERIA ZA9C-R38, Windows 10)

After targeting the eye pupil using a GUI, every image from the telephoto camera is processed within 2 ms so that the overall system runs at 500 Hz. Since the processing time for wide-angle camera images can exceed 2 ms, processing up to a maximum of approx. 4 ms is always performed on the latest image so that the delay in processing results utilized for distance estimation is within two frames.

D. Estimation of 3D Eye Position

Assuming that the eyeball exists at the center of the telephoto camera image and that the rotation centers of the mirrors coincide, as shown in Fig. 5, the 3D position of the eye $\mathbf{X}$ in the wide-angle camera coordinate can be derived from the corresponding point (u, v) on the image

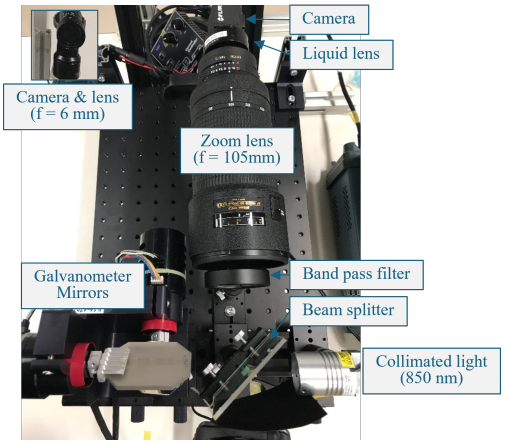

Fig. 4. Prototype system.

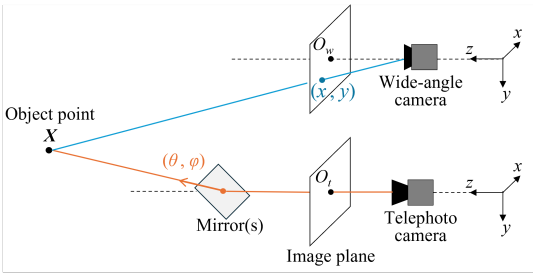

Fig. 5. Coordinate relationships among two cameras and gaze control mirrors under several assumptions. θ_m and ϕ_m represent horizontal and vertical angles of the gaze.

plane and the gaze angle (θ, ϕ) of the telephoto camera by using triangulation method. Suppose

$$w_c \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \boldsymbol{P}_c \boldsymbol{X}, \quad w_m \begin{pmatrix} \theta \\ \phi \\ 1 \end{pmatrix} = \boldsymbol{P}_m \boldsymbol{X} \ (\boldsymbol{P}_c, \boldsymbol{P}_m \in \mathbb{R}^{3\times 4}), \quad (1)$$

where w_c and w_m are scaling factors, $\boldsymbol{P}_c$ is a camera matrix of the wide-angle camera, $\boldsymbol{P}_m = [\boldsymbol{R}|-\boldsymbol{R}\boldsymbol{t}]$, $\boldsymbol{R}$ is a rotation matrix, and $\boldsymbol{t}$ is a translation vector from the mirror center, respectively. Since a cross product of two vectors in the same direction is zero, the following equations are derived from Eq. 1:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \times \boldsymbol{P}_c \boldsymbol{X} = \begin{pmatrix} v\boldsymbol{p}_{c3}^\top - \boldsymbol{p}_{c2}^\top \\ \boldsymbol{p}_{c1}^\top - u\boldsymbol{p}_{c3}^\top \\ u\boldsymbol{p}_{c2}^\top - v\boldsymbol{p}_{c1}^\top \end{pmatrix} \boldsymbol{X} = \boldsymbol{O},$$

$$\begin{pmatrix} \theta \\ \phi \\ 1 \end{pmatrix} \times \boldsymbol{P}_m \boldsymbol{X} = \begin{pmatrix} \phi\boldsymbol{p}_{m3}^\top - \boldsymbol{p}_{m2}^\top \\ \boldsymbol{p}_{m1}^\top - \theta\boldsymbol{p}_{m3}^\top \\ \theta\boldsymbol{p}_{m2}^\top - \phi\boldsymbol{p}_{m1}^\top \end{pmatrix} \boldsymbol{X} = \boldsymbol{O},$$

where $\boldsymbol{p}_n$ is the n-th row vector of matrix $\boldsymbol{P}$. There are four independent equations against $\boldsymbol{X}$ with three dimensions (the third line of each above is a linear combination of the first and second lines). As these equations are not equivalent due to the influence of noise, we adopted a simple linear approach called the linear-eigen method [18] to find a least squares solution $\boldsymbol{X}^*$ instead of finding the exact $\boldsymbol{X}$.

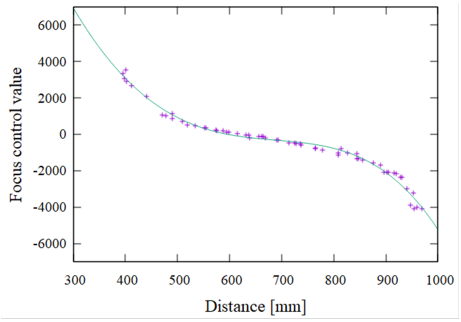

Fig. 6. Calibration results of relationship between depth distance of the target and liquid lens focus value.

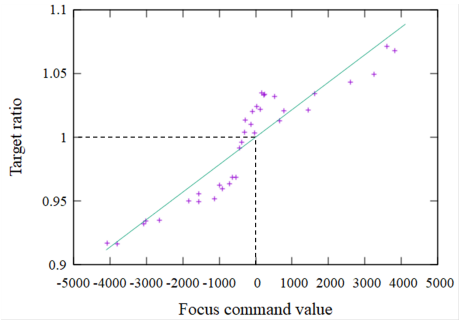

Fig. 7. Calibration results of relationship between focus distance and the focus control value for the liquid lens. We used the square root of the ellipse-approximation area size of the bright pupil as a measure of magnification (almost the same results as when using other measures such as pupil width).

E. Liquid Lens Focus Control

The instruction value to control the liquid lens focus can be determined by the depth distance of the target eye from the telephoto camera, with a known distance between the mirror center and the camera lens. Sueishi et al. [19] utilized a look-up table for determination, but as Wang et al. [20] noted that the relationship between the focus distance and the input voltage can be expressed as a curve, we approximated it using a 4th-order polynomial this time (Fig. 6).

F. Eye Angle Estimation

In the proposed system, several steps are required to convert the acquired pupil-CR distance into angle information.

The first is to compensate for image magnification that varies depending on the shooting distance of the eye and liquid lens control value. In an ordinary zoom lens, two mechanisms operate: one to change the focal length and the other to keep the image magnification constant. However, the proposed system only controls the focal length, so the image magnification changes by the liquid lens control value. Since the liquid lens control value is dependent on the shooting distance, if the frame rate is high enough, the magnification rate should be expressed only by the liquid lens control value. After checking the relationship between the liquid lens control value f and the magnification rate r in the prototype system (Fig. 7), we found that even a linear function can approximate with an error of less than 5%.

Next, to analyze the movement of the eyeball itself, the component associated with head motion should be removed

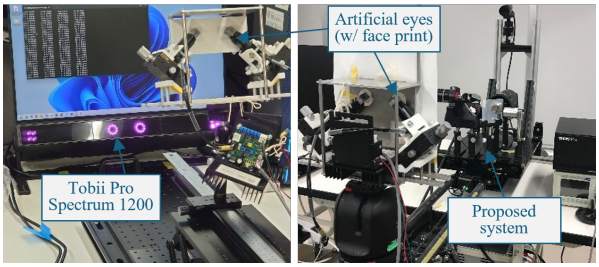

Fig. 8. Setup for evaluating Tobii Pro Spectrum (left) and the proposed system (right). Both capture the artificial eye [21] from slightly downward.

from the pupil-CR distance. Let $\boldsymbol{x}$ [px] be the pupil-CR distance with magnification correction that corresponds to the relative angle to the telephoto camera (referred to as "gaze angle"), then $\boldsymbol{x} = \boldsymbol{x}_e + \boldsymbol{x}_{ht} + \boldsymbol{x}_{hr}$, where $\boldsymbol{x}_e$, $\boldsymbol{x}_{ht}$ and $\boldsymbol{x}_{hr}$ represent the components associated with the eyeball rotation, head translation, and head rotation, respectively. $\boldsymbol{x}_{ht}$ is estimated from the gaze direction of the telephoto camera (i.e., the mirror angles) as follows:

$$\boldsymbol{x}_{ht} = k_1(\boldsymbol{a} \cdot \boldsymbol{\theta}_m + \boldsymbol{\theta_0}),$$

where k_1 is the conversion coefficient between pixel and angle, $\boldsymbol{\theta}_m$ is the mirror angle, and $\boldsymbol{\theta}_0$ is the origin angle of the eye at $\boldsymbol{\theta}_m = (0,0)^\top$. Each element of the slope $\boldsymbol{a}$ is ideally a constant value of 1, but is subject to calibration in consideration of disturbance effects in the mirror angle acquisition. The eyeball may shift slightly from the image center because the camera tracks not the eyeball itself but the pupil, but the resultant change in mirror angle is presumably negligible because the diameter of the eyeball is about 25 mm, which is sufficiently smaller than the shooting distance. The estimation of the head rotation component x_{hr} should be an issue for future work because we consider that binocular detection with a 3D face modeling is necessary to achieve sufficient accuracy. Then, the remaining $\boldsymbol{x}_e$ is multiplied by a fixed coefficient k_2 to obtain the eyeball rotation angle (referred to as "eye angle"). In this paper, k_1, k_2, $\boldsymbol{a}$ and $\boldsymbol{\theta}_0$ were calibrated using measurement data of two types of eye movements at different depth positions: a) with the eye stationary at the origin angle and b) with a square wave of amplitude of 1.0 deg generated with a 2-second cycle.

IV. PERFORMANCE EVALUATION

A. Setup

To validate the performance under unfixed and unrestrained head conditions, we compared the proposed system with a conventional device called Tobii (Tobii Pro Spectrum 1200 [10]), one of the most commonly used devices in microsaccade research, in terms of a) the measurement range and b) the robustness against head movements during measurement. Fig. 8 shows the experimental setup.

To ensure reproducibility of eye movements, we used a bright-pupil eye model [21] (referred to as "eye"), which allows fast and fine degree control. Since eye trackers need face information for measurements, we prepared a head model (referred to as "head") with a face print on top and

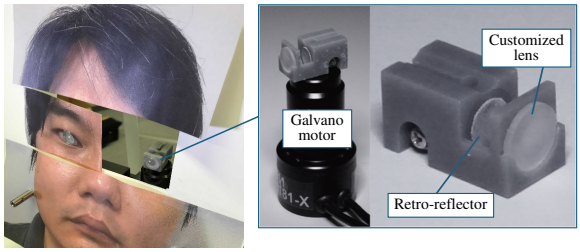

Fig. 9. Bright-pupil microsaccadic artificial eye [21] used in the experiment. A human face printed on paper was placed in front of the eye so that the trackers could detect face position.

TABLE I
MEASURABLE DISTANCE RANGE AGAINST STATIC AND MOVING HEAD

	Static head		Moving head	
	Tobii	Ours	Tobii	Ours
Lower limit [mm]	450	**300**	450	**300**
Upper limit [mm]	820	**950**	750	**950**
Measurable range [mm]	370	**650**	300	**650**

the eye model placed on the eye part, as shown in Fig. 9. The head was set on a linear slider in front of each eye tracker so that the head movement was limited to the depth direction, where the proposed device is expected to be the most effective compared to conventional eye trackers. In order to match the conditions as closely as possible, all measurements were made using the same calibration information. Tobii used a calibration result for 600 mm in the depth distance, and the proposed system calibrated the relationships between the cameras, mirrors, and liquid lens by using AR markers.

B. Measurement Range in Depth Direction

First, the measurement ranges in the depth direction were identified for each setup. Table I lists the results of measurement ranges where data can be acquired when the head is a) stationary and b) moving at a speed of 10 mm/s.

For Tobii, the measurement range for a static head was 450–820 mm, but not beyond 750 mm when the head moved. This is presumably because of a focal blur at the far end and even a small amount of motion blur caused a loss of CR in the image. In contrast, the proposed system covered a measurement range of 300–950 mm for both static and moving heads, nearly double that of Tobii, with a sufficiently low noise level.

C. Data Quality

Next, we examined precision and accuracy, which are typical indicators utilized to evaluate eye tracker performance, using data acquired at each measurable distance per 5-second period of time. The precision is a measure of spatial resolution calculated as the root mean square (RMS) and standard deviation (SD) of inter-sample distances at eye rest, which is desired to be less than 0.05 deg for detecting fine movements to a minimum of 0.2 deg. The accuracy was calculated as the error in the amplitude estimate of the square

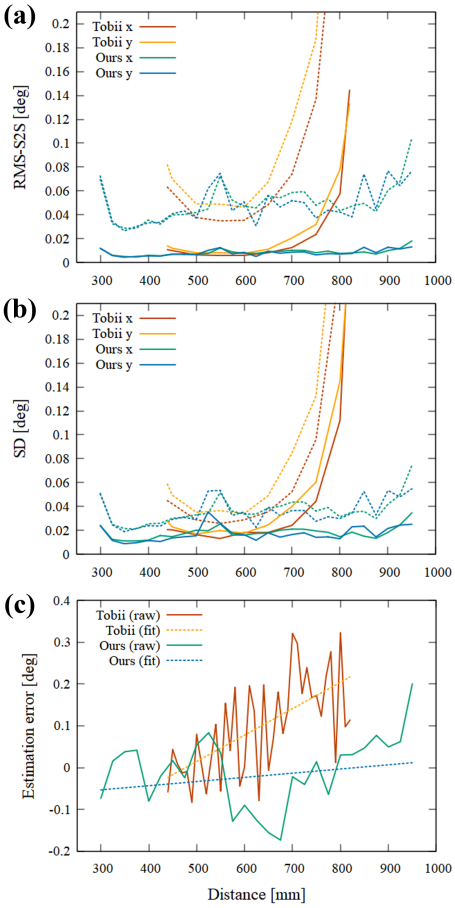

Fig. 10. (a) Root mean square (RMS) of sample-to-sample distances. (b) Standard deviation (SD) of the gaze signal. (c) Amplitude estimation errors as accuracy scores. (a)–(c) contain results of each recording setup: Tobii Pro Spectrum 1200 (Tobii) and the proposed system (Ours). "x" and "y" refer to the horizontal and vertical dimensions of the data.

wave motion of a known amplitude of 1.0 deg. Fig. 10 shows the results under each recording setup.

For the precision, the dotted and solid lines respectively denote the results for raw data and data smoothed with a Savitzky-Golay filter of 21 ms, where the filter size was set based on the discussion in [10] so that high-frequency noise could be removed without destroying the fine eye movements. Tobii showed stable accuracy within a range of approx. 300 mm wide centered at the calibrated distance of 600 mm, but outside of this range the quality deteriorated noticeably, or the eye was out of the angle of view. In contrast, the proposed device achieved sufficient precision of less than 0.05deg over the entire measurement range of 650 mm.

In the accuracy plot, Tobii tends to estimate larger amplitudes at greater distances, indicating that it is difficult to guarantee the eye angle as the depth position of the head changes from the calibration point. The proposed system shows a smaller difference in estimation trends by distance and appears to maintain a constant error range.

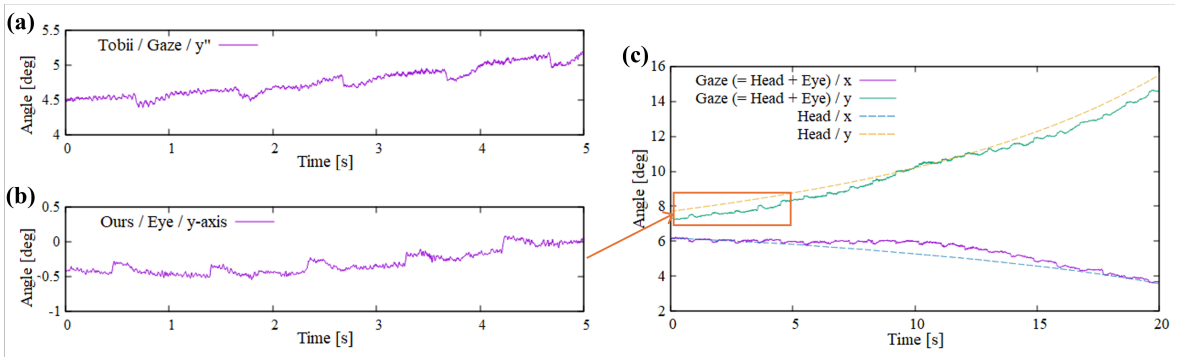

Fig. 11. (a) Vertical gaze angles obtained by Tobii Pro Spectrum 1200. (b) Vertical eye angles (i.e., gaze angles minus head translation components) obtained by the proposed system. (c) Gaze angles obtained by the proposed system and the estimated head translation components. The head moved forward at an approximate speed of 20 mm/s in both cases, and the eye generated microsaccadic movements of 0.2 deg every second.

D. Robustness against Head Movement

To investigate robustness against head movements, data were acquired for each setup by moving the head on the linear slider at a speed of 10 to 40 mm/s while the eye generated saccadic movements of 0.2 deg every second. The maximum speed is determined by the limitation of the linear slider. Fig. 11 shows the results in a 5-second period of time when the head was moved forward at an approximate speed of 20 mm/s in each setup. All data were filtered by a Savitzky-Golay filter of 21 ms. Both Tobii and the proposed system were able to track the eye stably with small noise in the entire measurement range. Tobii was expected to be more robust against very fast head movements because of its high frame rate, but no superior differences were observed at the speed of up to 40 mm/s.

E. Head Movement Range in Sitting Position

Finally, we demonstrated the actual range of the head movement as one of the authors sat while peering into the screen and leaning back. The eye level from the seat height was 770 mm and the head moved over a range of 340 mm, which can be covered only by the proposed system. The angle of inclination was then approx. $\pm$12.5 deg. For details on the estimated depth distances during this demonstration, see the experimental results in the next chapter (Fig. 15).

Although difficult to discuss due to the author's lack of medical expertise, it is interesting that microsaccadic movements occur even while the head moves from back to front. It might be significant to be able to measure such events.

V. STUDY FOR PRACTICAL USE

A. Dataset for Microsaccade Detection

The experimental results discussed in the previous section demonstrate that the proposed system allows for a wide measurement range of 650 mm in depth direction while acquiring visible data down to a microsaccade of 0.2 deg. For the next step, we studied a policy to detect saccades and microsaccades automatically using the proposed system. A list of the data utilized in this study is provided in Table II.

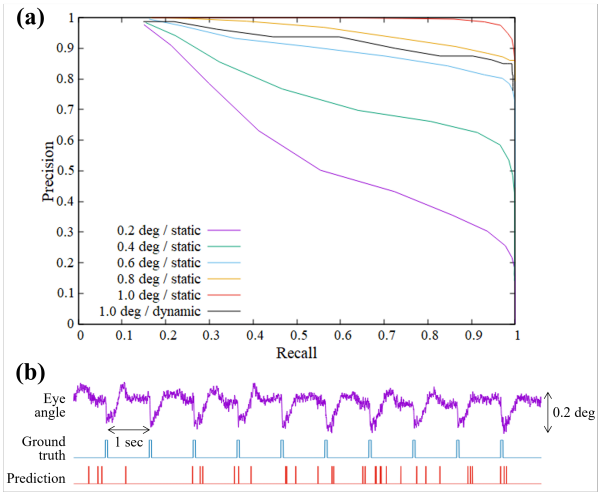

Fig. 12. (a) PR curve of detection results based on the thresholding approach [22]. Labels indicate the microsaccade amplitude and the head movement status. (b) Example of a detection result with low accuracy (amplitude 0.2 deg / distance 600 mm).

In data 1 - 3, the eye generated microsaccadic movements of a predetermined amplitude (0.2 - 1.0 deg) every second. Data 4 measured the actual eye movements of one of the author.

B. Rule-based Method

The most common approach to microsaccade detection is to implement peak-velocity-based thresholding [22]. This method allows setting a dynamic threshold based on the standard deviation of each measurement data. However, since microsaccades correlate with amplitude and peak velocity ([23]), the finer the motion, the smaller the peak velocity, and the more likely it is to be confused with high-frequency noise. Fig. 12 shows the PR curves and an example of the prediction result when this method is applied to Data 1 and 2. The accuracy is noticeably worse for small amplitudes, especially those below 0.4 deg. Moreover, the accuracy for Data 2 seems to be adversely affected by the velocity offset due to the estimation error of the head movement component.

TABLE II

DATA UTILIZED IN STUDY OF MICROSACCADE DETECTION

#	Eye type	Head	Distance	Amplitude	Total data length
1	Artificial	Static	300–950 mm	0.2–1.0 deg	350 s
2	Artificial	Dynamic (uniform linear motion)	300–950 mm	0.2 deg, 1.0 deg	192 s
3	Artificial	Dynamic (figure "8" motion)	600–740 mm	0.2–1.0 deg	100 s
4	Human	Dynamic (free motion)	600–940 mm	—	20 s

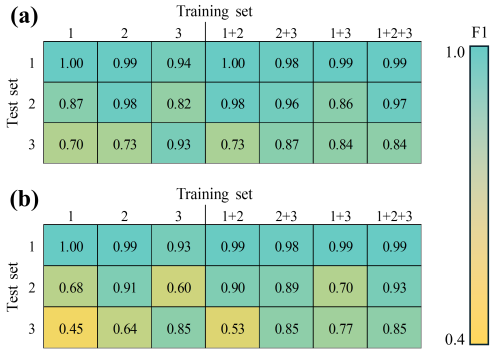

Fig. 13. The best F1 values for different combinations of training and test sets. (a) Results for eye angle data (i.e., gaze angles minus head translation components). (b) Results for gaze angle data that contains head translation components. For each pattern, we averaged five results with different seed values for data split and learning.

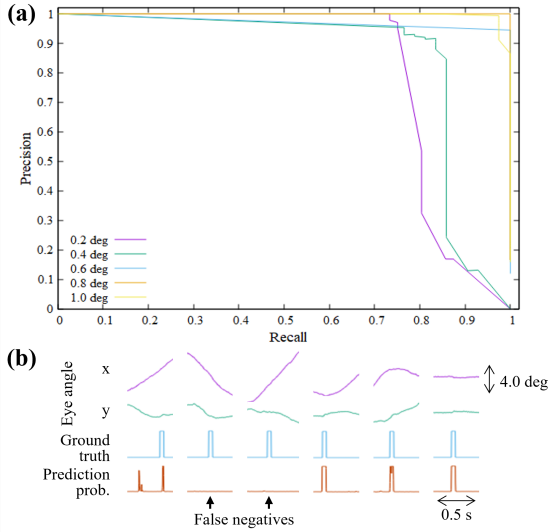

Fig. 14. (a) PR curve by amplitude under a training set of 1+2+3 and a test set of 3. (b) Fragments of detection results with amplitude of 0.2 deg.

C. CNN-based Method

In light of the above, we focused on U'n'Eye [24], a method that classifies motion type with a CNN-based discriminator, as an approach that can detect fine motion and has high flexibility for head movement. We retained the default model parameters and used differential waveforms (i.e., velocities) as inputs to exclude the effects of the gaze angle offset. The data length per input was set to 500 frames (equivalent to 1 second), as 1 second or longer was found to be good through trials. The duration of each microsaccade in Data 1–3 was defined as 25 frames, roughly three times wider than the visually confirmed duration, considering the latency between the trigger input to the eye model and the waveform observation. In addition, to suppress false positives, fires with a continuous firing interval of less than six frames were rejected.

Fig. 13 shows the prediction performance (best F1 values, harmonic mean of recall and precision) for different combinations of training and test sets. Notably, using data with the head translation component removed (Fig. 13 (a)) improved the performance of most patterns compared to data with the component remaining (Fig. 13 (b)). This means the removal worked appropriately so that the differences in motion properties among data were canceled to a certain extent. Training sets that included both data 2 and 3 yielded good overall results, but test set 3 showed a decrease in recall for microsaccades with amplitudes below 0.4 deg (Fig. 14). A detailed check revealed that there were few false positives in general, and that intervals with a high change rate in the eye angle tended to result in false negatives. Assuming that the head translation component was properly removed, the head rotated quickly in those intervals. It is rather surprising that microsaccades with amplitudes over 0.6 deg were visible even during head rotation. There are a few possible reasons for the false negatives, such as that data 1 and 2 did not contain head rotations or that data 2 was too small to cover all variations of motion. However, since microsaccades typically occur during prolonged visual fixation, it is unlikely that they occur during dynamic head rotation, and this result is thus considered sufficient for practical use.

Furthermore, we applied the model trained on data 1+2+3 to data 4 (Fig. 15). Although we lack medical knowledge to conduct rigorous analysis, the model fired appropriately for all visible saccadic movements with a minimum jumping of approx. 0.2 deg, indicating that we achieved sufficient accuracy even if the training set contained only eye model data. This finding will be helpful in future experiments with humans or other biological life form.

The overall results indicated that the proposed system and the CNN-based detector achieved high accuracy in detecting microsaccades of down to 0.2 deg (0.6 deg with head rotation), even when the head moves dynamically.

VI. CONCLUSION

In this paper, we proposed a desktop eye tracker that can measure microsaccadic eye movements without requiring a chin rest. The proposed eye tracker system composed of two cameras, rotating mirrors, a liquid lens, and an infrared

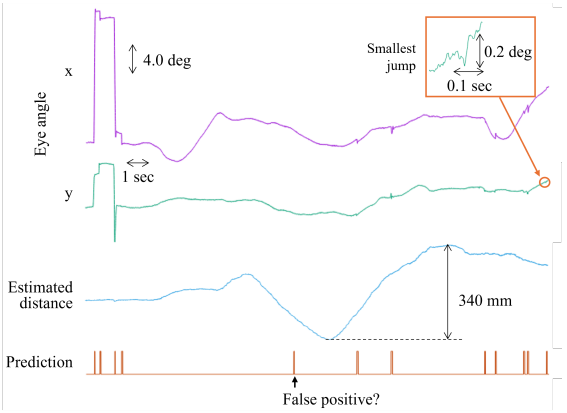

Fig. 15. Eye angles, depth distance, and saccade detection results of data 4 with a model trained on data 1+2+3.

light controlled in real-time at 500 Hz. We investigated the usefulness of the proposed system through comparative experiments with an existing high-accuracy desktop eye tracker [10] and confirmed that the proposed system had a wider measurement range in depth with sufficient temporal and spatial resolution to detect microsaccades of down to 0.2 deg. We also examined automatic microsaccade detection, confirming that the CNN-based method [24] could detect small microsaccades even when the head moved dynamically and that the model trained with eye model data could be applied to human eye data.

We consider these results as the first step toward the practical application of high-resolution eye trackers such as to monitor driver concentration based on microsaccade frequency and to measure psychological state of a subject during counseling without him/her being aware of it. By making the system smaller and increasing the shooting distance, it would be possible to monitor driver concentration based on microsaccade frequency and to measure the psychological state of a subject during counseling without his/her being aware of it. Future work should include conducting experiments in real-world scenarios against human subjects to confirm that the proposed system really works. The neural network had better be improved to separate the head movements internally. Neurophysiological studies using the proposed system are also necessary to find further possible applications, and it becomes important to show the relationship between the movements of the right and left eyes and to separate the gaze angle into rotation components of the eyeball and the head. Therefore, we will also plan to develop a binocular system that combines one wide-angle camera and two telephoto cameras.

REFERENCES

[1] P. M. Tostado, W. W. Abbott, and A. A. Faisal, "3d gaze cursor: Continuous calibration and end-point grasp control of robotic actuators," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 3295–3300.

[2] I. Tong, O. Mohareri, S. Tatasurya, C. Hennessey, and S. Salcudean, "A retrofit eye gaze tracker for the da vinci and its integration in task execution using the da vinci research kit," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015, pp. 2043–2050.

[3] X. Gao, H. Yan, and H.-j. Sun, "Modulation of microsaccade rate by task difficulty revealed through between- and within-trial comparisons," *Journal of Vision*, vol. 15, no. 3, pp. 3–3, 03 2015.

[4] J. Takahashi, Y. Hirano, K. Miura, K. Morita, M. Fujimoto, H. Yamamori, Y. Yasuda, N. Kudo, E. Shishido, K. Okazaki, T. Shiino, T. Nakao, K. Kasai, R. Hashimoto, and T. Onitsuka, "Eye movement abnormalities in major depressive disorder," *Front. Psychiatry*, vol. 12, Aug. 2021.

[5] R. G. Alexander, S. L. Macknik, and S. Martinez-Conde, "Microsaccade characteristics in neurological and ophthalmic disease," *Front. Neurol.*, vol. 9, Mar. 2018.

[6] Z. Kapoula, Q. Yang, J. Otero-Millan, S. Xiao, S. L. Macknik, A. Lang, M. Verny, and S. Martinez-Conde, "Distinctive features of microsaccades in alzheimer's disease and in mild cognitive impairment." *Age*, vol. 36, no. 2, pp. 535–543, Sep. 2013.

[7] E. D. Guestrin and M. Eizenman, "General theory of remote gaze estimation using the pupil center and corneal reflections," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 6, pp. 1124–1133, 2006.

[8] K. Holmqvist and P. Blignaut, "Small eye movements cannot be reliably measured by video-based p-cr eye-trackers," *Behavior Research Methods*, vol. 52, no. 5, pp. 2098–2121, Oct 2020. [Online]. Available: https://doi.org/10.3758/s13428-020-01363-x

[9] [Online]. Available: https://www.sr-research.com/eyelink-1000-plus/

[10] M. Nyström, D. C. Niehorster, R. Andersson, and I. Hooge, "The tobii pro spectrum: A useful tool for studying microsaccades?" *Behavior Research Methods*, vol. 53, pp. 335–353, 2021.

[11] "Metric data 02 average dimensions of person sitting." [Online]. Available: https://www.firstinarchitecture.co.uk/metric-data-02-average-dimensions-of-person-sitting/

[12] D. H. Yoo and M. J. Chung, "A novel non-intrusive eye gaze estimation using cross-ratio under large head motion," *Comput. Vis. Image Underst.*, vol. 98, no. 1, p. 25 – 51, apr 2005.

[13] S. Kim, M. Sked, and Q. Ji, "Non-intrusive eye gaze tracking under natural head movements," in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, 2004, pp. 2271–2274.

[14] E. Hampsey, P. G. Overton, and T. Stafford, "Microsaccade rate as a measure of drug response," *J Eye Mov Res*, vol. 12, no. 6, Oct. 2019.

[15] R. G. Alexander, A. Venkatakrishnan, J. Chanovas, S. L. Macknik, and S. Martinez-Conde, "Microsaccades mediate perceptual alternations in monet's "impression, sunrise"," *Scientific Reports*, vol. 11, no. 1, p. 3612, Feb. 2021.

[16] A. Sato, H. Imaoka, T. Suzuki, and T. Hosoi, "Advances in face detection and recognition technologies," 2005.

[17] I. Ishii, Y. Nakabo, and M. Ishikawa, "Target tracking algorithm for 1 ms visual feedback system using massively parallel processing," *Proceedings of IEEE International Conference on Robotics and Automation*, vol. 3, pp. 2309–2314 vol.3, 1996. [Online]. Available: https://api.semanticscholar.org/CorpusID:46261242

[18] R. I. Hartley and P. Sturm, "Triangulation," *Computer Vision and Image Understanding*, vol. 68, no. 2, pp. 146–157, 1997.

[19] T. Sueishi, T. Ogawa, S. Yachida, and M. Ishikawa, "Continuous high-resolution observation system using high-speed gaze and focus control with wide-angle triangulation," in *High-Speed Biomedical Imaging and Spectroscopy V*, K. K. Tsia and K. Goda, Eds., vol. 11250, International Society for Optics and Photonics. SPIE, 2020, p. 1125012.

[20] L. Wang, Y. Hu, H. Xu, and M. Ishikawa, "Dynamic focal tracker display," in *Advances in Display Technologies IX*, J.-H. Lee, Q.-H. Wang, and T.-H. Yoon, Eds., vol. 10942, International Society for Optics and Photonics. SPIE, 2019, p. 109420K.

[21] T. Sueishi, S. Matsumura, S. Yachida, and M. Ishikawa, "Optical and control design of bright-pupil microsaccadic artificial eye," *In Proc. of IEEE/SICE Intl. Symposium on System Integration (SII)*, pp. 760–765, 2022.

[22] R. Engbert and R. Kliegl, "Microsaccades uncover the orientation of covert attention," *Vision research*, vol. 43, no. 9, pp. 1035–1045, Apr. 2003.

[23] B. L. Zuber, L. Stark, and G. Cook, "Microsaccades and the velocity-amplitude relationship for saccadic eye movements," *Science*, vol. 150, no. 3702, pp. 1459–1460, Dec. 1965.

[24] M. E. Bellet, J. Bellet, H. Nienborg, Z. M. Hafed, and P. Berens, "Human-level saccade detection performance using deep neural networks," *Journal of Neurophysicology*, vol. 121, no. 2, pp. 646–661, Feb. 2019.