

Optical Engineering

OpticalEngineering.SPIEDigitalLibrary.org

Discriminating between intentional and unintentional gaze fixation using multimodal-based fuzzy logic algorithm for gaze tracking system with NIR camera sensor

Rizwan Ali Naqvi
Kang Ryoung Park

SPIE.

Rizwan Ali Naqvi, Kang Ryoung Park, "Discriminating between intentional and unintentional gaze fixation using multimodal-based fuzzy logic algorithm for gaze tracking system with NIR camera sensor," *Opt. Eng.* **55**(6), 063109 (2016), doi: 10.1117/1.OE.55.6.063109.

Discriminating between intentional and unintentional gaze fixation using multimodal-based fuzzy logic algorithm for gaze tracking system with NIR camera sensor

Rizwan Ali Naqvi and Kang Ryoung Park*

Dongguk University, Division of Electronics and Electrical Engineering, 30, Pildong-ro 1-gil, Jung-gu, Seoul 100-715, Korea

Abstract. Gaze tracking systems are widely used in human–computer interfaces, interfaces for the disabled, game interfaces, and for controlling home appliances. Most studies on gaze detection have focused on enhancing its accuracy, whereas few have considered the discrimination of intentional gaze fixation (looking at a target to activate or select it) from unintentional fixation while using gaze detection systems. Previous research methods based on the use of a keyboard or mouse button, eye blinking, and the dwell time of gaze position have various limitations. Therefore, we propose a method for discriminating between intentional and unintentional gaze fixation using a multimodal fuzzy logic algorithm applied to a gaze tracking system with a near-infrared camera sensor. Experimental results show that the proposed method outperforms the conventional method for determining gaze fixation. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.OE.55.6.063109](https://doi.org/10.1117/1.OE.55.6.063109)]

Keywords: accuracy of gaze detection; discriminating between intentional and unintentional gaze fixation; gaze tracking system using near-infrared camera sensor.

Paper 160541 received Apr. 12, 2016; accepted for publication Jun. 10, 2016; published online Jun. 23, 2016.

1 Introduction

The field of human–computer interaction (HCI) has grown significantly over recent decades. Bolt¹ described how eye-gaze information could be used as the input for facilitating HCI, and patterns of eye movements and fixations have been found to be usable indicants of the distribution of visual attention and important indicants of thinking processes.² The use of gaze input to trigger computer operations is becoming increasingly popular. The idea of using computer-assistive technology for interaction with personal computers (PCs) via devices such as switches, head pointers, neural interfaces, and eye-tracking systems was proposed by Mauri et al.³ As these devices require activation by body parts, they often cannot be used by severely disabled people who cannot control their hands, feet, or head. Some devices used by disabled people can be controlled using bioelectrical signals or a switch.⁴ Physiological information, such as that from electroencephalograms (EEGs), electromyograms (EMGs), and electroculograms (EOGs), provides an alternative communication method for patients with severe motor disabilities.⁵ Using EEG signals, i.e., based on brain waves, people can control screen keyboards, mice, or wheelchairs.^{5,6} EMG bioelectrical signals based on muscle response can be used to interact with other systems,^{5,7} and EOG signals can be used for simple interaction purposes because they determine the approximate gaze direction based on eye movement.^{5,8,9} Devices used for measuring bioelectrical signals are expensive and can irritate the subject because sensors must be placed on the body. Hence, camera-based gaze detection methods are preferred as the alternative.

The two-dimensional (2-D) monitors of desktop computers have been used in eye-gaze tracking methods.^{10–12} However, these methods have some limitations, e.g., they cannot control the devices in the three-dimensional (3-D) space, and their accuracy worsens when there are variations in the Z-distances between the user and the monitor. Therefore, nonwearable gaze tracking systems for controlling home appliances in the 3-D space have been proposed.¹³ Most studies on gaze detection have focused on enhancing the accuracy of gaze detection, whereas few have considered the discrimination between intentional gaze fixation (looking at a target to activate or select it) and unintentional fixation while using gaze detection systems. A user's gaze fixation can be classified as visually motivated (unintentional) fixation (looking at something to see it) and interaction motivated (intentional) fixation (looking at something to activate or select it). In this study, we focus on interaction motivated (intentional) fixation.

To discriminate between different types of gaze fixation, researchers have used methods based on keyboard or mouse button clicking, eye blinking, and the dwell time of gaze position. However, these techniques are limited in terms of user convenience, selection speed, and so on.

Previous studies on gaze fixation can be categorized into those that use single or multiple modalities to select the object of interest. The former category^{14–21,22–27} includes eye blinks, dwell time, antisaccades, “on” and “off” screen buttons, context switching, keystrokes, eyebrow raises, and speech. Blinking to select letters from the alphabet is an obvious solution for eye typing when the gaze direction is used to select letters.¹⁴ However, eye blinks normally occur at a rate of ~10/min,¹⁵ and it would be necessary to close the eye for a longer period to discriminate between eye blinking for letter selection and normal blinking, which decreases user

*Address all correspondence to: Kang Ryoung Park, E-mail: parkgr@dgu.edu

convenience. Object selection based on dwell time appears more natural than selection by blinking.¹⁶ For this, the gaze tracking system has to be conscious of where the user is looking and of how long he/she looks at an object in order to select it.

Hansen et al.¹⁷ used the dwell time of the user gaze position for letter selection in an eye typing application, and Hornhof and Cavender¹⁸ proposed a system in which various menus within a drawing program can be selected using the dwell time of the user gaze position. Huckauf and Urbina¹⁹ developed a target selection approach that uses antisaccades rather than blink selection or dwell time. Antisaccades are explicit eye movements that have been extensively examined in cognitive psychology.²⁰ Ware and Mikaelian²¹ used on and off buttons for object selection. In their method, an object of interest is selected by fixation and subsequent saccade toward the on/off buttons.

In previous studies,¹⁴⁻²⁷ the object of interest is pointed at and selected by a single modality, but such methods suffer from the problem whereby objects become selected every time the user looks at them. This limitation was first referred to as the “Midas Touch Problem.”²⁸ “Midas Touch Problem” is from Greek mythology, which tells that even the objects which King Midas does not want to select are transformed into Gold. A similar case occurs in gaze tracking system. That is, the case that a user is looking at the object of interest with intention (selecting or activating it) should be discriminated from that the user is looking at without any intention. This is “Midas Touch Problem” in gaze detection system.

To overcome this problem, the object of interest should be discriminated from those objects that are unintentionally fixated. However, when object selection is performed by blinking, it is difficult to discriminate between intentional and unintentional blinks. Selection by dwell time encounters similar issues, i.e., if the dwell time is too long, it can tire the user’s eyes and result in slower task performance,^{17,18} whereas if the dwell time is too short, we encounter the Midas Touch Problem. Graphical on/off screen buttons can be problematic, because they interfere with the relevant object and distract the user from the area or object of interest. Zhai et al.²² and Kumar et al.²³ combined gaze control with manual input, i.e., keystrokes, for pointing at and selecting

objects of interest. Grauman et al.²⁴ proposed a method based on blinking or raising an eyebrow to point at and select objects and convey commands. Kaur et al.²⁵ proposed the idea of complementing gaze control with speech. Surakka et al.²⁶ suggested the idea of frowning to select the object of interest. Tuisku et al.²⁷ proposed a text entry method that relies on gazing and smiling, where gaze is used to point at an object and smiling is used as the selection tool. However, these techniques do not satisfy the requirements of patients with severe motor disabilities, e.g., amyotrophic lateral sclerosis patients who cannot move any part of their body except the eyes. To overcome the limitations of single modality-based methods, this study examines a multimodal approach based on pupil accommodation and a short dwell time.

In previous research, Verney et al.^{29,30} indicated that cognitive tasks can affect changes in pupil size. Based on this, we adopt the spontaneous change of pupil size (pupil accommodation) as one modality for analyzing the fixation and nonfixation of user gaze for near-infrared (NIR) camera-based gaze tracking systems. The proposed approach is unique in four ways:

- First, we propose the use of pupil accommodation as an indicator for the fixation and nonfixation of gaze position in actual gaze tracking systems.
- Second, the concept of peakedness is introduced to measure pupil accommodation with respect to time.
- Third, we use the features of the change in pupil size (for measuring pupil accommodation) and change in gaze position over a short dwell time to investigate gaze fixation and nonfixation phenomena.
- Fourth, a fuzzy system is adopted using these two features as inputs, and the gaze fixation or nonfixation decision is made through defuzzification.

Table 1 gives a comparative summary of the proposed and existing methods. The main distinction between the proposed and existing methods is that object of interest is selected by one modality (single modality-based methods) or plural modalities (multiple modalities-based methods). As single modality-based methods, there exist the methods based on eye blink,^{14,24} dwell time,^{17,18} antisaccades,^{19,20} on and off screen buttons,²¹ keystrokes,^{22,23} eyebrow raises,²⁴

Table 1 Comparison between previous and proposed methods for object selection.

Category	Single modality	Multiple modalities
Method	Object of interest is selected by eye blink, ^{14,24} dwell time, ^{17,18} antisaccades, ^{19,20} on and off screen buttons, ²¹ keystrokes, ^{22,23} eyebrow raises, ²⁴ speech, ²⁵ face frowning, ²⁶ and smiling ²⁷	Pupil accommodation with short dwell time by fuzzy system (proposed method)
Advantages	-Simpler implementation than multimodal methods -Less input data to be processed than in multimodal methods	Higher accuracy of object selection (the success rate of target selection) compared with single modality methods
Disadvantages	-Accuracy of object selection (the success rate of target selection) is lower than multimodal methods -Small errors in the sensing data can adversely affect the overall results -Some methods are difficult to use for patients with severe motor disabilities, especially those who can only move their eyes ²²⁻²⁷	Necessity of designing a fuzzy rule table and membership function

speech,²⁵ face frowning,²⁶ and smiling.²⁷ For example, in case of the method based on eye blink, the object of interest on a screen can be selected after a user's gazing at it and his (or her) eye blinking being perceived by the method. In case of the method based on dwell time, the object of interest can be selected after a user's gazing at it and the maintenance of the status of gazing (for predetermined time period) being perceived. Our method belongs to multiple modalities-based methods because two modalities such as pupil accommodation and short dwell time are checked for user's selecting the object of interest. For example, in our method, the object of interest can be selected after user's gazing at it, and both pupil accommodation and the maintenance of the status of gazing (for short time period) being perceived.

The remainder of this paper is organized as follows. The proposed system and methodology are introduced in Sec. 2. In Sec. 3, the experimental setup is described and the results are presented. Section 4 draws together our conclusions and discusses some ideas for future work.

2 Object Selection by Pupil Accommodation with Short Dwell Time

2.1 Overview of Proposed Method

In the proposed method, a commercial web camera (Logitech C600³¹) with universal serial bus interface and NIR

illuminator (wavelength of 850 nm) of 8×8 NIR light-emitting diodes (LEDs) are used for the eye-tracking device. Illumination by NIR LEDs can reduce glare to a user's eye and distinguish the boundary between the pupil and iris in an eye image.³² In detail, with the NIR light of shorter wavelength [700 (or 750) to 800 nm], the iris becomes darker (compared to the case using the NIR light of 850 nm), which causes the reduction of distinctiveness of boundary between the pupil and iris in the image. Therefore, it is more difficult to detect the correct pupil area in the image. With the NIR light of longer wavelength (higher than 900 nm), the iris becomes brighter (compared to the case using the NIR light of 850 nm), which causes the increase of distinctiveness of boundary between the pupil and iris in the image. Therefore, it is easier to locate the correct pupil area in the image. However, the sensitivity of camera sensor generally decreases according to the increase of the wavelength of illuminator. Therefore, the captured image by the NIR light higher than 900 nm becomes so dark that correct detection of pupil area is difficult. Therefore, we use the NIR illuminator of 850 nm in our gaze tracking system.

The image resolution for the eye-tracking camera is set to 1600×1200 pixels to obtain more accurate gaze estimation. Our system captures images at a rate of 15 frames per second (fps). An NIR-passing filter is used to ensure that the images captured by the eye-tracking camera are not affected by

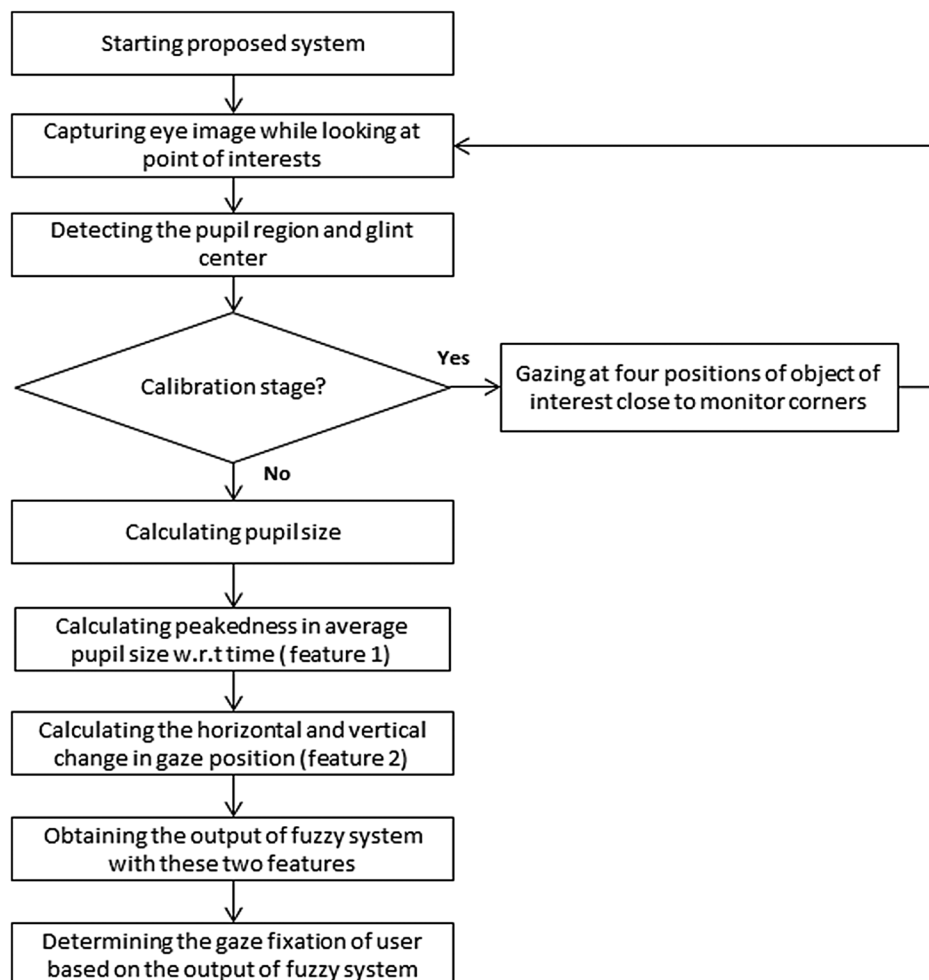


Fig. 1 Overall procedure for the proposed method.

exterior visible light conditions. The eye-tracking camera is equipped with a zoom lens to obtain large eye images. Although various commercial gaze tracking systems are available,^{33–36} they do not provide any functionality for measuring the change of pupil size. As this is needed in our system to determine gaze fixation, we constructed a bespoke gaze tracking system.

A flowchart for the proposed system is shown in Fig. 1. Our gaze-tracking camera first acquires images of the user's eye while the user is looking at objects of interest. From the captured eye image, the glint center and pupil region are located (see details in Sec. 2.2). Here, glint refers to the bright spot on the corneal surface caused by the NIR illuminator. A user-dependent calibration is then performed while the user gazes at the four positions of the object of interest. These positions are close to the corners of the monitor. After the user calibration step, the pupil size is measured based on the major and minor axes calculated by pupil ellipse fitting (see details in Sec. 2.3). To measure the pupil accommodation, peakedness is calculated based on the average pupil size with respect to time as feature 1 (F_1) (see details in Sec. 2.3). The change of gaze position in the horizontal and vertical directions is then calculated over a short dwell time as feature 2 (F_2) (see details in Sec. 2.4). From F_1 and F_2 , we calculate the output value of the fuzzy system. Subsequently, the fixation and nonfixation of user gaze are determined based on the fuzzy output, and, in the case of fixation, the object of interest is selected (see details in Sec. 2.5).

2.2 Preprocessing Steps for Detection of Pupil and Glint Centers

Our system locates the pupil region and glint center. A flowchart for this procedure is shown in Fig. 2. This flowchart corresponds to the “Detecting the pupil region and glint center” step of Fig. 1. With the captured eye image, the glint candidates are extracted using image binarization, labeling, and size-based filtering methods in the predefined search region. If the glint exists in the search region, the region of interest (ROI) is defined based on the glint candidate, and the approximate region of the pupil is detected in the ROI using a sub-block-based matching method. This method defines nine sub-blocks, and the position of maximum difference between the mean of the gray level of the central sub-block (block 4 in Fig. 3) and those of the surrounding sub-blocks (0 to 3 and 5 to 8 in Fig. 3) is determined as the approximate pupil region. To enhance the processing speed of the sub-block-based method, an integral imaging method is adopted when calculating the average intensity of each sub-block.^{37,38}

The reason why the pupil region is detected within the searching area defined by the located glint is that the position of glint is usually close to that of the pupil. As shown in Fig. 17, the NIR illuminator is close to our gaze tracking camera, and the camera is also close to the monitor, which the user is looking at. Therefore, the position of glint produced by the NIR illuminator is close to that of pupil in the captured image.

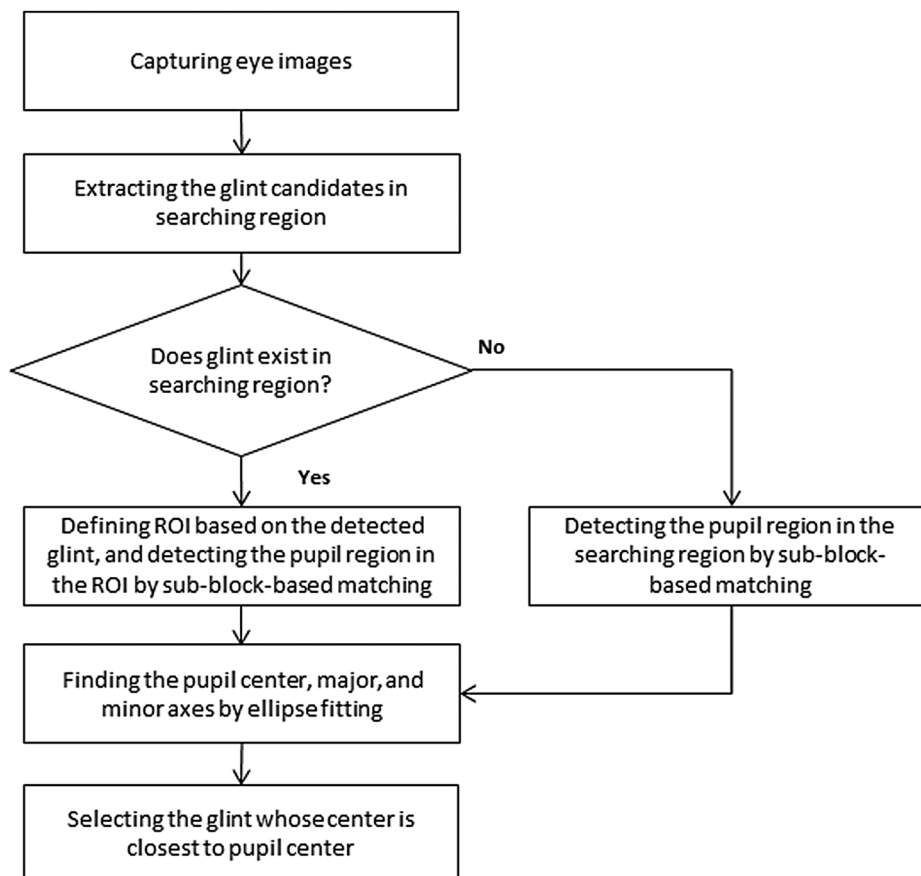


Fig. 2 Flowchart for detection of glint center and pupil region.

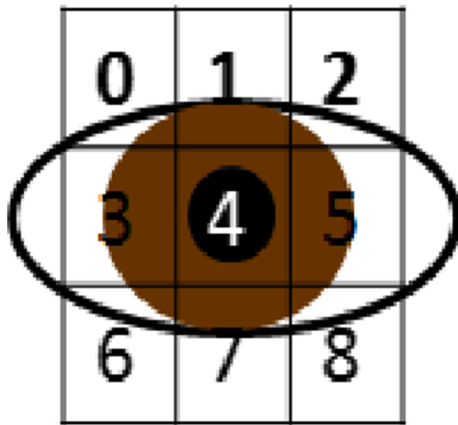


Fig. 3 Mask of sub-block-based matching for pupil detection.

If there is no glint in the search region, the sub-block-based matching method is performed in the search region to detect the approximate pupil region. The size of the sub-blocks varies from 20×20 to 60×60 pixels to cope with pupils of different sizes. Figure 4 shows the detection of the glint and the approximate pupil region. The whole eye region is divided into two parts, and the sub-block-based matching method is performed in each part.

Within the approximate pupil region, the accurate pupil center and the major and minor axes of the pupil region are detected by ellipse fitting (as shown in Fig. 5), and the glint whose center is closest to the pupil center is selected.¹³

The process for detecting the accurate pupil center is shown in Fig. 5. This flowchart corresponds to the “Finding the pupil center, major, and minor axes by ellipse fitting” step of Fig. 2. Within the approximate pupil region in

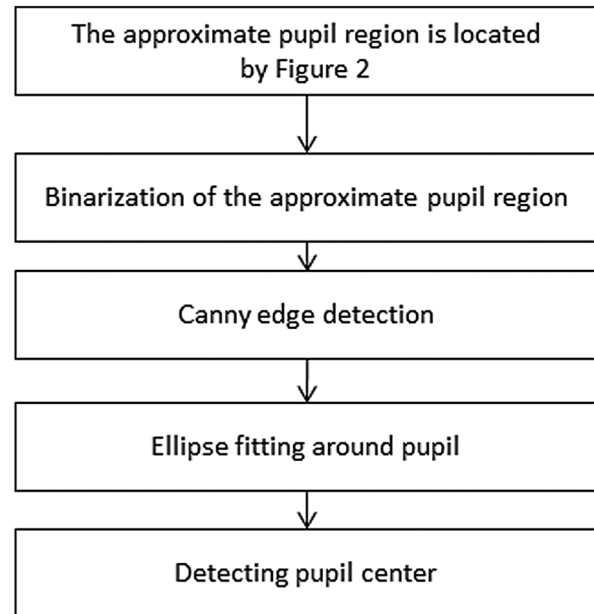


Fig. 5 Flowchart for detection of pupil center.

Fig. 6(a), histogram stretching is performed to increase the distinction between the pupil and iris areas, as shown in Fig. 6(b). Image binarization is then performed, as shown in Fig. 6(c), using the threshold value determined by Gonzalez’s method.³⁹ The boundary of the pupil region is found using a Canny edge detector,⁴⁰ as shown in Fig. 6(d). As it can be seen in Fig. 6(e), ellipse fitting is used to find the pupil area. The major and minor axes of ellipse fitting can then be obtained as shown in Fig. 6(f), and the final result of the pupil center and boundary detection is given in Fig. 6(g).

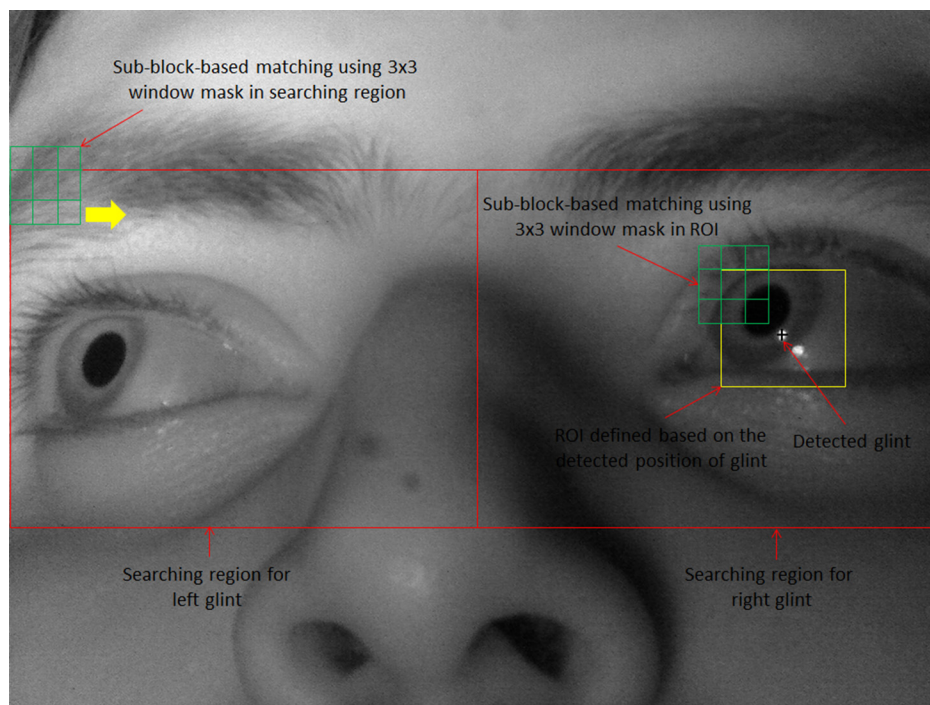


Fig. 4 Example of detecting glint and approximate pupil region. Box on left eye shows case where glint is not located, whereas that on right eye represents case where glint is located successfully.

2.3 Calculating Peakedness in Average Pupil Size with Respect to Time as Feature 1

Figure 7 shows changes in the size of the pupil while the user is looking at an object of interest. In the image, the pupil size can be calculated by fitting an ellipse around the pupil boundary and determining the major and minor axes of the ellipse,

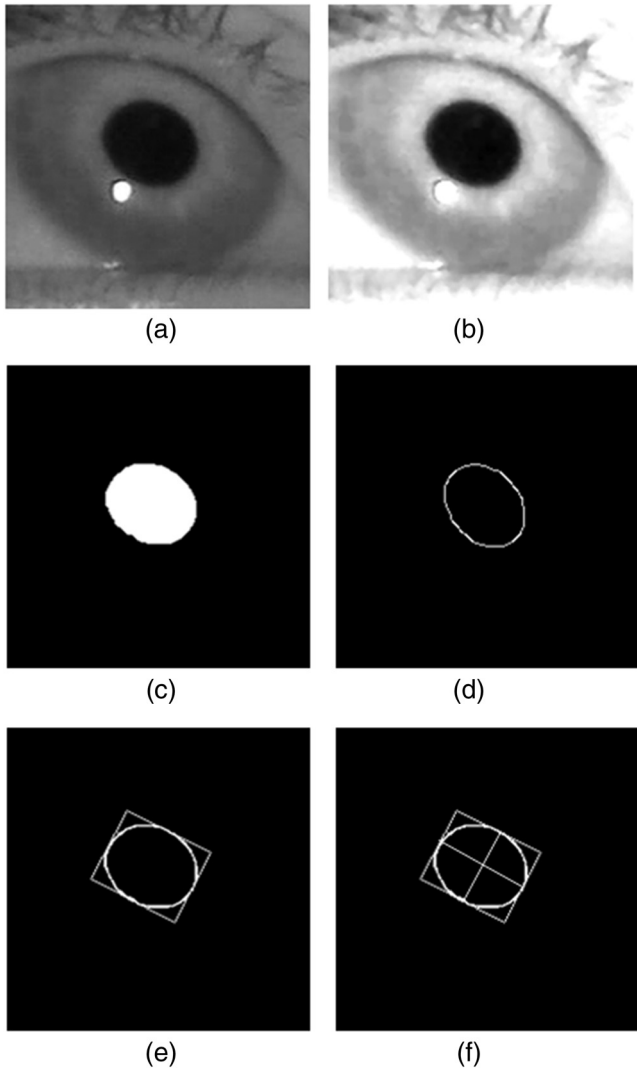


Fig. 6 Procedure for accurately detecting pupil centers. (a) Original image. (b) Histogram stretched image. (c) Binarized image. (d) Result by canny edge detection. (e) Ellipse fitting with canny edge image. (f) Image with major and minor axes of ellipse fitting. (g) Final result of detected pupil center and boundary.

as shown in Fig. 8. Equation (1) is used to calculate the size of the pupil.⁴¹

$$\text{Size of Ellipse (Pupil)} = \pi \times a \times b. \tag{1}$$

Based on Eq. (1), we can obtain a graph of the change in pupil size with respect to time. A moving average filter based on three coefficients (1/3, 1/3, and 1/3) can then be applied to the graph to reduce noise.⁴² Using the filtered graph, the gradient of the average pupil size can be obtained.

We set up a camera to capture eye images at 15 fps. Therefore, the time required for each frame is 1/15 s (66.6 ms). It has been experimentally observed that the maximum time required by the pupil to constrict and dilate is less than 600 ms. Based on this, we use a window of 10 frames to observe pupil dilation and constriction. Using this window, the peakedness (Pk) is calculated as

$$\text{Peakedness}(Pk) = \left(\sum_{i=P}^{D-1} g'_i \right) - \left(\sum_{i=D}^{W+P-1} g'_i \right), \tag{2}$$

where g'_i is the gradient between two adjacent points on the graph of the change in pupil size with respect to time. D is the time of the peak on the graph, and W is the size of the window, i.e., 10 frames. P is the estimated start (time)

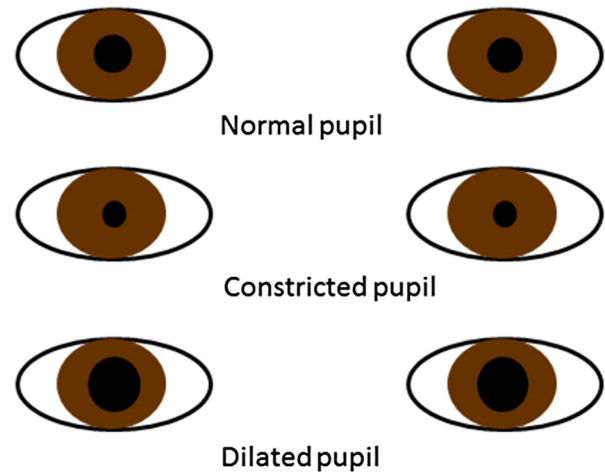


Fig. 7 Example of variations in pupil size while looking at object of interest.

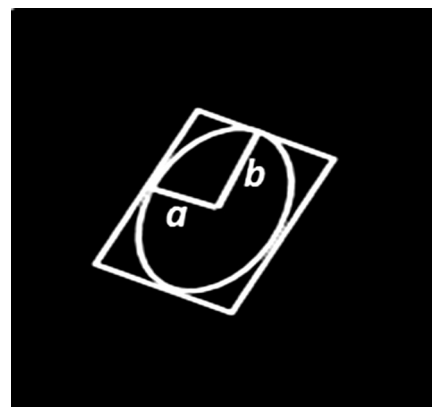


Fig. 8 Example of fitted ellipse with major and minor axes.

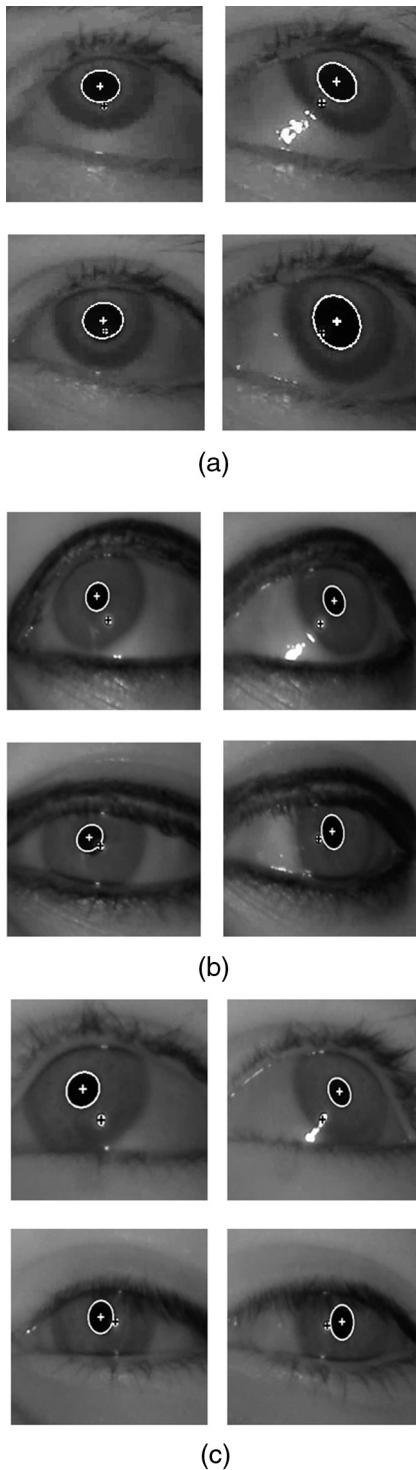


Fig. 9 Examples of four images including the detected centers of pupil and glint when a user is looking at the four calibration positions on monitor. (a) Example 1, (b) example 2, and (c) example 3. In (a)–(c), the upper-left, upper-right, lower-left, and lower-right figures show the cases that each user is looking at the upper-left, upper-right, lower-left, and lower-right calibration positions on monitor, respectively.

position of gaze fixation. In our research, D is determined by checking the gradient of the graph of changes in pupil size. The position of P is detected based on changes in gaze position in the horizontal and vertical directions. Based on previous results showing that cognitive tasks can affect the

change in pupil size,^{29,30} we expect Pk to increase with the large change in pupil size in the case of gaze fixation. By subtracting Pk from its maximum value (determined from experimental data in advance), the smaller value of Pk indicates gaze fixation.

In order to reduce the measured error, we use the average value of pupil accommodation [Pk of Eq. (2)] of both eyes. In the case of gaze fixation, pupil size is first increased and then decreased. In order to measure this phenomenon in the captured successive images, the graph of the change in pupil size with respect to time is measured in the image as shown in Fig. 11(a). As observed in Fig. 11(a), the pupil size (blue line) changes (first increases, then decreases) after the starting (time) positions (red line) of gaze fixation.

Because the pupil size is measured by the size of the ellipse of the pupil in the image as shown in Eq. (1), the unit of the pupil size is pixels. From the graph like Fig. 11(a), the gradient (g_i) between two adjacent points on the graph are measured. That is, the gradient (g_i) is the difference of pupil size in two adjacent points, and the unit of the difference of pupil size is pixels. Because the two adjacent points are obtained from two successive images and our system captures images at a rate of 15 fps, the time interval between two adjacent points is 66.7 ms (1000/15). Consequently, the gradient (g_i) is the difference (pixels) of pupil size in two successive images per time (66.7 ms) between two successive images. Therefore, the unit of gradient (g_i) is pixels/(66.7 ms). By multiplying 66.7 to the original measured gradient (g_i), we obtain the revised gradient (g'_i) (whose unit is pixels/ms). This revised gradient (g'_i) is summated for time period as shown in Eq. (2). Therefore, peakedness (Pk) means the sum of pupil size change within a time window [W of Eq. (2)], and its unit is also pixels/ms. Peakedness (Pk) represents the magnitude of pupil state changes. The peakedness (feature 1) and changes in gaze position (feature 2, which is explained in Sec. 2.4) are normalized to be in the range of 0 to 1 before being used as the two inputs to fuzzy system. Therefore, the multiplication by 66.7 and the difference of unit between feature 1 (pixels/ms) and feature 2 (pixel) do not affect the performance of our system.

2.4 Calculating Horizontal and Vertical Changes in Gaze Position within Short Dwell Time as Feature 2

To obtain feature 2, the gaze position is calculated based on the detected pupil center and glint center (explained in Sec. 2.2).¹³ To calculate the gaze position, each user looks at four positions close to the monitor corners during the initial calibration stage, and we obtain four pairs of pupil centers and glint centers, as shown in Fig. 9.

With these four pairs of detected pupil centers and glint centers, the position of the pupil center is compensated based on the glint center to reduce the variation in gaze position caused by head movements. With these four pairs of detected pupil centers and glint centers, a geometric transform matrix can be calculated. This matrix defines the relationship between the pupil movable region and the monitor region, as shown in Fig. 10. In general, the relationship of transformation between two quadrangles can be defined by multiple unknown parameters.⁴³ If the transformation just includes in-plane rotation and translations (on x - and y -axes), the

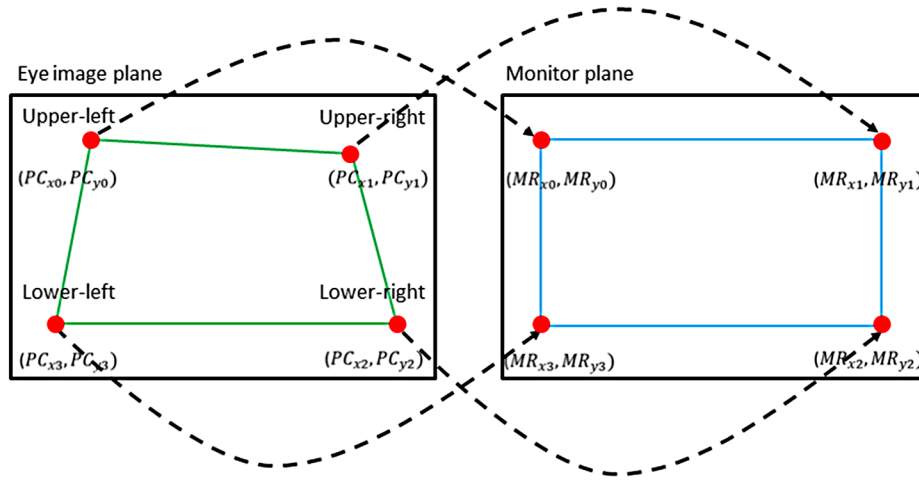


Fig. 10 Relationship between the pupil movable region [the quadrangle defined by (PC_{x0}, PC_{y0}) , (PC_{x1}, PC_{y1}) , (PC_{x2}, PC_{y2}) , and (PC_{x3}, PC_{y3})] on the eye image and the monitor region [the quadrangle defined by (MR_{x0}, MR_{y0}) , (MR_{x1}, MR_{y1}) , (MR_{x2}, MR_{y2}) , and (MR_{x3}, MR_{y3})].

relationship can be defined using three unknown parameters (Euclidean transform). If the transformation includes in-plane rotation, translations (on x - and y -axes), and scaling, the relationship can be defined using four unknown parameters (similarity transform). In the case that the transformation includes in-plane rotation, translations (on x - and y -axes), scaling, and parallel skewing, the relationship can be defined using six unknown parameters (affine transform). As the last case, if the transformation includes in-plane rotation, translations (on x - and y -axes), scaling, parallel skewing, and distortion, the relationship can be defined using eight unknown parameters (projective or geometric transform). In our research, we consider the last case for defining the relationship of transformation between two quadrangles of the pupil movable region and the monitor region, by which various transform can be covered by our method. Therefore, we use eight unknown parameters in Eq. (3).

This geometric transform matrix is calculated by Eq. (3), and the user's gaze position (GP_x, GP_y) is given by Eq. (4).¹³

$$\begin{bmatrix} MR_{x0} & MR_{x1} & MR_{x2} & MR_{x3} \\ MR_{y0} & MR_{y1} & MR_{y2} & MR_{y3} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} a & b & c & d \\ e & f & g & h \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} PC_{x0} & PC_{x1} & PC_{x2} & PC_{x3} \\ PC_{y0} & PC_{y1} & PC_{y2} & PC_{y3} \\ PC_{x0}PC_{y0} & PC_{x1}PC_{y1} & PC_{x2}PC_{y2} & PC_{x3}PC_{y3} \\ 1 & 1 & 1 & 1 \end{bmatrix}, \tag{3}$$

$$\begin{bmatrix} GP_x \\ GP_y \end{bmatrix} = \begin{bmatrix} a & b & c & d \\ e & f & g & h \end{bmatrix} \begin{bmatrix} PC'_x \\ PC'_y \\ PC'_xPC'_y \\ 1 \end{bmatrix}. \tag{4}$$

Because the left and right eyes usually gaze at the same position, we obtain the gaze positions of the left and right eyes, and use the average of the two positions as the final gaze position. Based on the user's gaze position, feature 2 is calculated by taking the sum of the differences in horizontal or vertical change in gaze position between the current and previous frames. The absolute value of the sum is taken from the estimated start (time) position of gaze fixation [P in Eq. (2)] over a short dwell time [the time window size W in Eq. (2)]. The absolute values of the sum in the X and Y directions are called AVSX and AVSY, respectively,

$$AVSX = \sum_{i=P}^{W+P-1} |\Delta x_i|, \tag{5}$$

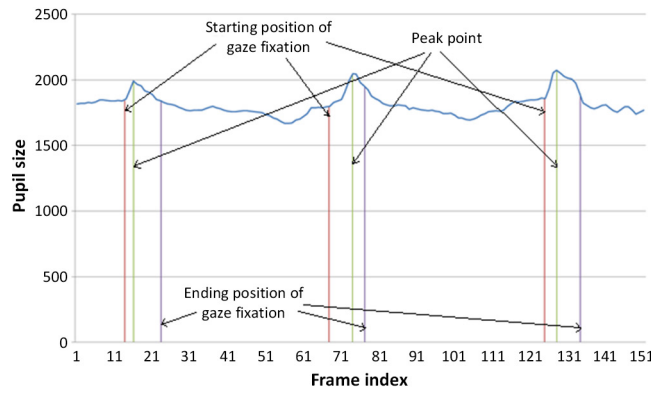
$$AVSY = \sum_{i=P}^{W+P-1} |\Delta y_i|. \tag{6}$$

The final value of feature 2 [change in gaze (Δd)] is determined by selecting the larger of AVSX and AVSY

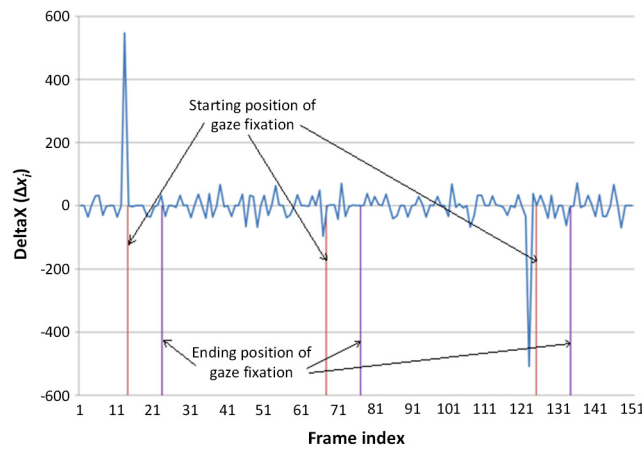
$$\text{Change in gaze}(\Delta d) = \begin{cases} AVSX & \text{if } (AVSX \geq AVSY) \\ AVSY & \text{else if } (AVSX) \end{cases}. \tag{7}$$

We expect the change in gaze (Δd) of Eq. (7) to be smaller in the case of user fixation, because the differences (Δx or Δy) in the horizontal or vertical changes in gaze position between the current and previous frames will be smaller. As shown in Eqs. (5)–(7), among the two summations (AVSX and AVSY) of absolute value of Δx and Δy , the larger one is selected as the change of gaze (Δd) (feature 2). Therefore, no threshold is used for this procedure, and the unit of the change of gaze (Δd) is also pixel. Then, the change of gaze (Δd) is normalized as 0 to 1, and it is used as one input (feature 2) to fuzzy system. Therefore, no threshold is required.

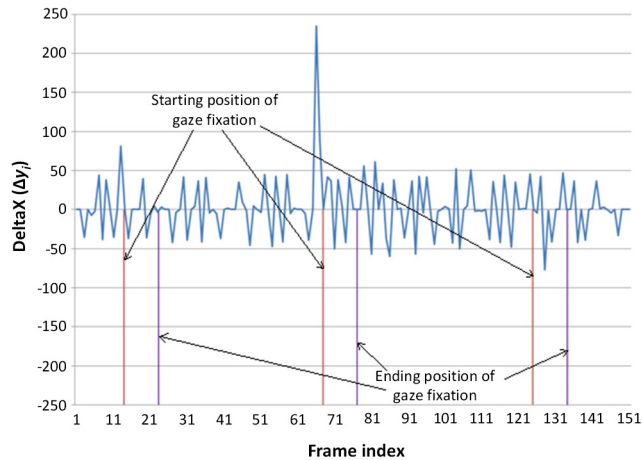
An example of the graph for the variation in pupil size with respect to time is shown as a blue line in Fig. 11(a). In addition, the graphs for Δx_i and Δy_i [from Eqs. (5) and (6)]



(a)



(b)



(c)

Fig. 11 Variations in pupil size and Δx_i and Δy_i [from Eqs. (5) and (6)] with respect to time. (a) Variations in pupil size. (b) Δx_i . (c) Δy_i .

with respect to time are shown as blue lines in Figs. 11(b) and 11(c), respectively. In Figs. 11(a)–11(c), the estimated start (time) positions of gaze fixation [P from Eqs. (2), (5), and (6)] and the end positions of gaze fixation ($W + P - 1$ from Eqs. (2), (5), and (6)) are shown as red and violet lines, respectively. In addition, the detected peak on the graph [D from Eq. (2)] is shown as a green line in Fig. 11(a).

As observed in Fig. 11, the pupil size changes after the start (time) positions of gaze fixation. In addition, the change of gaze in the horizontal or vertical directions [Δx_i and Δy_i from Eqs. (5) and (6)] becomes smaller after the start (time) positions of gaze fixation. From this, we can confirm that these two features [peakedness (Pk) and change in gaze (Δd)] can be used as inputs to the fuzzy system for determining user gaze fixation.

2.5 Determination of Gaze Fixation Based on Fuzzy Logic System

2.5.1 Definition of fuzzy membership functions

To determine gaze fixation, the proposed method uses a fuzzy logic system with Pk and Δd as inputs, as shown in Fig. 12. As explained in Secs. 2.3 and 2.4, these two features decrease in the case of user gaze fixation. Through normalization based on minimum-maximum (min-max) scaling, Pk and Δd range from 0 to 1. Based on the output value of the fuzzy logic system, we can determine whether user gaze fixation has occurred.

Figure 13 shows the input membership functions for the input values of Pk and Δd . The input values are categorized into three groups in the membership function: low (L), medium (M), or high (H). In general, these three value groups are not separated. The membership functions are defined as the overlapped area shown in Fig. 13. Considering the processing speed and complexity of the problem to be solved, we use a linear (triangular) membership function, which has been widely adopted in fuzzy-based applications.⁴⁴⁻⁴⁶ The shapes of three input membership functions generally represent the overall distribution of input data as three functions (L, M, and H). In the fuzzy system, the shapes of these functions are not generally determined by training with data, but heuristically defined by human expert.

The input values are converted into degrees of membership using these membership functions. To determine

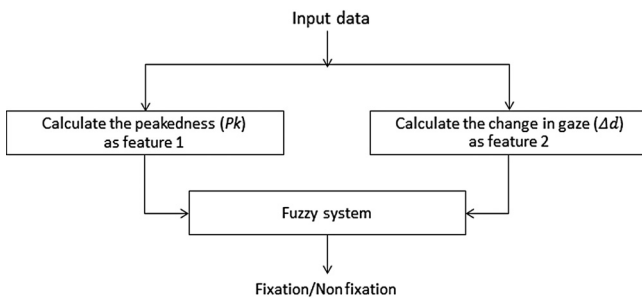


Fig. 12 Fuzzy-based method for determination of gaze fixation and nonfixation.

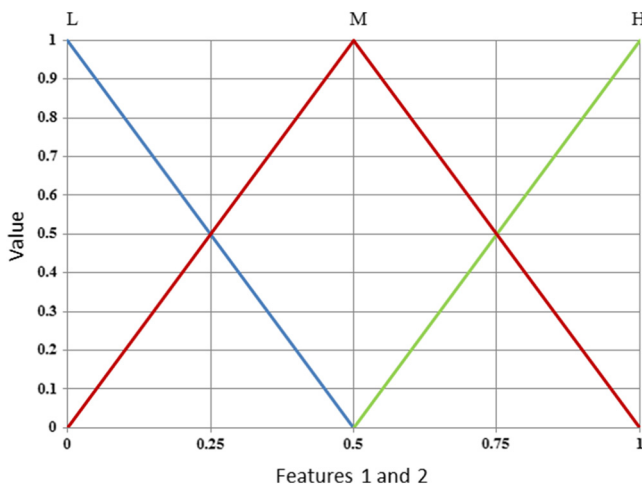


Fig. 13 Input fuzzy membership functions for input feature 1 [peakedness (Pk)] and input feature 2 [change in gaze (Δd)].

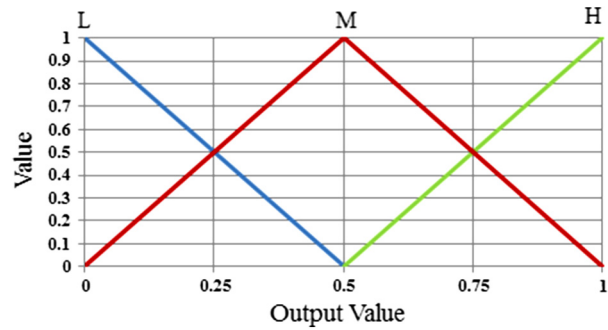


Fig. 14 Definition of output membership functions.

whether gaze fixation has occurred, the membership function for the output value is also defined in the form of a linear function (Fig. 14) that includes the three groups of L, M, and H. Using these output membership functions, the optimal output value is obtained from the defuzzification rule and membership degrees, which are explained in Sec. 2.5.2.

2.5.2 Fuzzy rules based on two input values

As explained in Secs. 2.3 and 2.4, both Pk and Δd become smaller in the case of user's gaze fixation. In this situation, we expect the probability of gaze fixation to be high (H). Therefore, we use the following rule of "if Pk and Δd are L, then output of fuzzy system becomes H" in Table 2. In addition, as these two features decrease when there is no gaze fixation, we can expect the probability of gaze fixation to be low (L) in this case. Based on these observations, we define the fuzzy rules listed in Table 2.

2.5.3 Determination of gaze fixation using defuzzification method

Using the two normalized input values, the corresponding six values can be obtained using the input membership functions shown in Fig. 15. We define the two membership functions as $f_{Pk}(\cdot)$ and $f_{\Delta d}(\cdot)$. The corresponding output values of the two functions with input values of Pk and Δd are denoted as $f_{Pk}^L, f_{Pk}^M, f_{Pk}^H, f_{\Delta d}^L, f_{\Delta d}^M, f_{\Delta d}^H$. For example, suppose that the two input values for Pk and Δd are 0.30 and

Table 2 Fuzzy rules based on Pk and Δd .

Pk	Δd	Output of fuzzy system
L	L	H
L	M	H
L	H	M
M	L	H
M	M	M
M	H	L
H	L	M
H	M	L
H	H	L

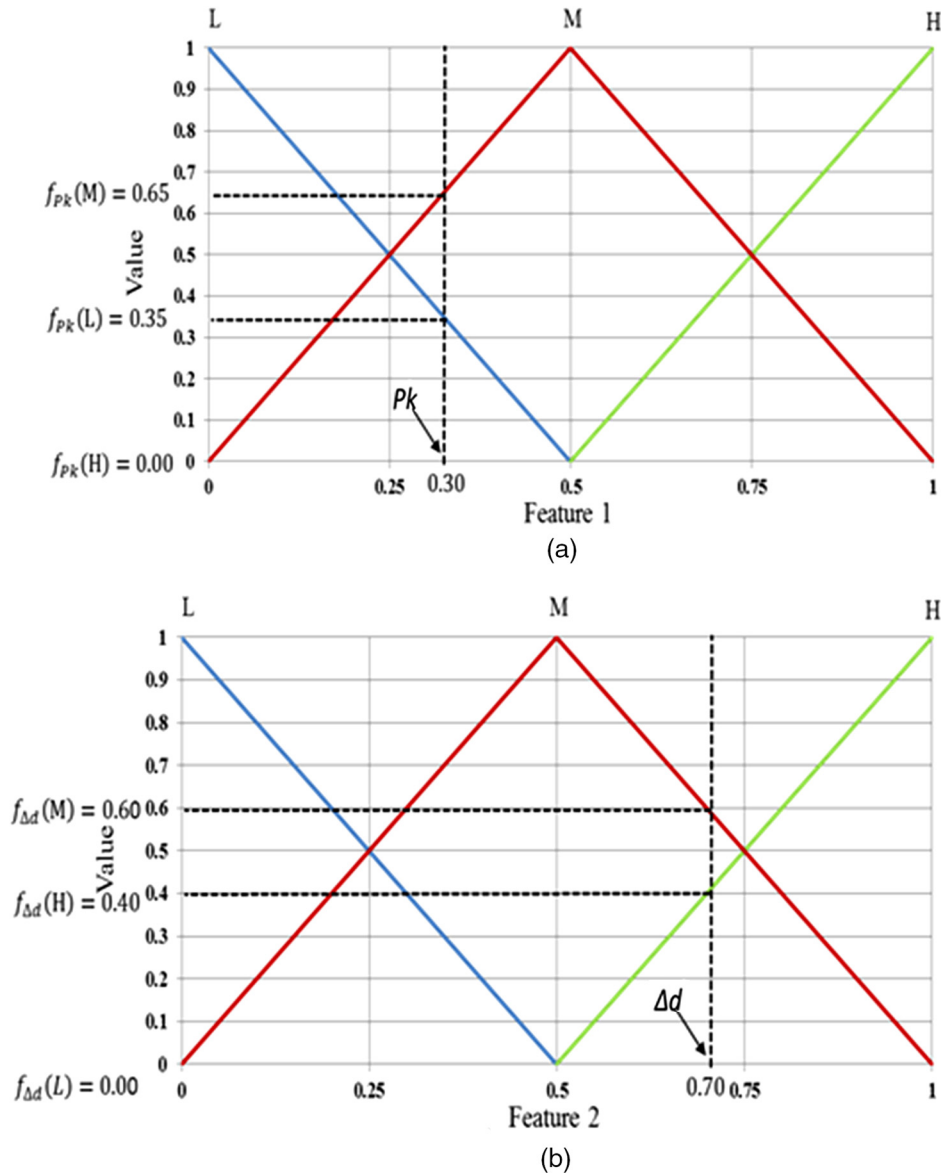


Fig. 15 Obtaining output value of input membership function for two features: (a) P_k and (b) Δd .

0.70, respectively, as shown in Fig. 15. The values of $f_{P_k}^L$, $f_{P_k}^M$, $f_{P_k}^H$, $f_{\Delta d}^L$, $f_{\Delta d}^M$, and $f_{\Delta d}^H$ are 0.35 (L), 0.65 (M), 0.00 (H), 0.00 (L), 0.60 (M), and 0.40 (H), respectively, as shown in Fig. 15. With these values, we can obtain the following nine combinations: [(0.35 (L), 0.00 (L)), [0.35 (L), 0.60 (M)], [0.35 (L), 0.40 (H)], [0.65 (M), 0.00 (L)], [0.65 (M), 0.60 (M)], [0.65 (M), 0.40 (H)], [0.00 (H), 0.00 (L)], [0.00 (H), 0.60 (M)], and [0.00 (H), 0.40 (H)].

With the nine fuzzy rules in Table 2, the proposed method determines which of L, M, and H can be used as the input for the defuzzification step. For this purpose, the MIN or MAX method is commonly used. In the MIN rule method, the minimum value is selected from each combination pair and used as the input for defuzzification. For the MAX rule, the maximum value is selected and used as the input for defuzzification. For example, for a combination pair of [0.35 (L), 0.60 (M)], the MIN rule selects the minimum value (0.35) as the input. For the MAX rule, the maximum value (0.60) is selected. Based on the fuzzy logic rules from Table 2 (if L

and M, then H), the value of 0.35 (H) and 0.60 (H) are finally determined by the MIN and MAX rules, respectively.

In Table 3, we list all of the values calculated by the MIN or MAX rules with the nine combinations {[0.35 (L), 0.00 (L)], [0.35 (L), 0.60 (M)], [0.35 (L), 0.40 (H)], [0.65 (M), 0.00 (L)], [0.65 (M), 0.60 (M)], [0.65 (M), 0.40 (H)], [0.00 (H), 0.00 (L)], [0.00 (H), 0.60 (M)], and [0.00 (H), 0.40 (H)]}. In this research, we refer to these values as the “inference values” (IVs). As indicated in Table 3, these IVs are used as the inputs for defuzzification in order to obtain the output. In our experiments, the MIN and MAX rules are compared.

Figure 16 shows several defuzzification methods used in our research. We consider five such methods: first of maxima (FOM), last of maxima (LOM), middle of maxima (MOM), mean of maxima (MeOM), and center of gravity (COG).^{45–48} In each method, the maximum IVs are used to calculate the output value. Figures 16(a) and 16(b) show that the maximum IVs are $IV_1(H)$ and $IV_2(M)$. Using the FOM method,

Table 3 IVs obtained with nine combinations.

Output of feature 1	Output of feature 2	IV	
		MIN rule	MAX rule
0.35 (L)	0.00 (L)	0.00 (H)	0.35 (H)
0.35 (L)	0.60 (M)	0.35 (H)	0.60 (H)
0.35 (L)	0.40 (H)	0.35 (M)	0.40 (M)
0.65 (M)	0.00 (L)	0.00 (H)	0.65 (H)
0.65 (M)	0.60 (M)	0.60 (M)	0.65 (M)
0.65 (M)	0.40 (H)	0.40 (L)	0.65 (L)
0.00 (H)	0.00 (L)	0.00 (M)	0.00 (M)
0.00 (H)	0.60 (M)	0.00 (L)	0.60 (L)
0.00 (H)	0.40 (H)	0.00 (L)	0.40 (L)

the first defuzzification value is selected as the optimal score value and represented as s_2 in Fig. 16(a). The LOM method selects the last defuzzification value as the optimal score value, i.e., s_4 . In the MOM method, the optimal score value is calculated using the average of the values obtained by FOM and LOM. Therefore, the output score value obtained by MOM is $[s_{MOM} = (1/2)(s_2 + s_4)]$. The MeOM method selects the mean of all defuzzification values as the output score value. The final output score value obtained by the MeOM method is calculated by $[s_{MeOM} = (1/3)(s_2 + s_3 + s_4)]$.

The output score in the COG method is calculated differently from the other defuzzification methods. The COG method calculates the output score value based on the geometrical center of the nonoverlapped area formed by the regions defined by all IVs. As it is shown in Fig. 16(b), the areas R_1 , R_2 , and R_3 are defined based on all IVs. R_1 is the quadrangle defined by the four points $[0, IV_3(L)]$, $[s_1, IV_3(L)]$, $(0.5, 0)$, and $(0, 0)$. R_2 is the quadrangle defined by the four points $[s_2, IV_2(M)]$, $[s_3, IV_2(M)]$, $(1, 0)$, and

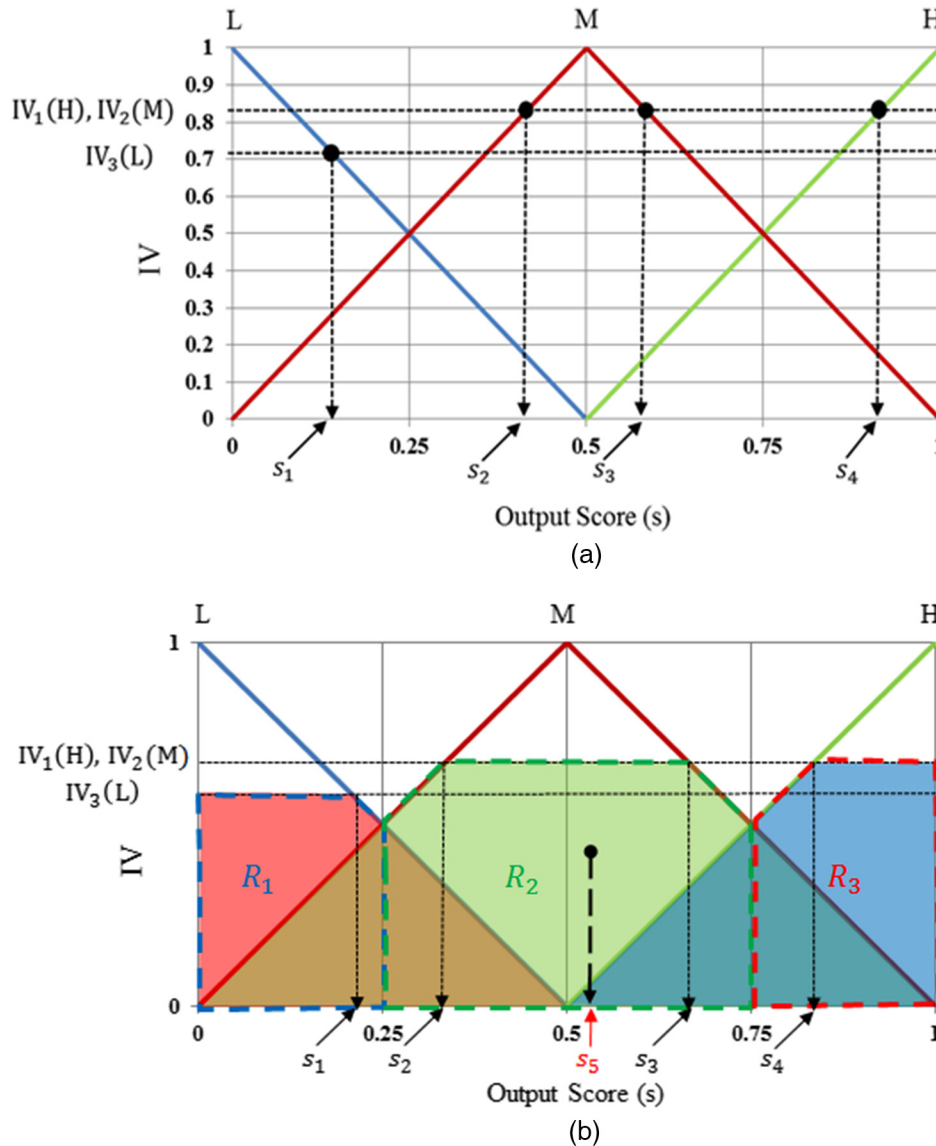


Fig. 16 Obtaining output score values of fuzzy system using different defuzzification methods. (a) FOM, LOM, MOM, and MeOM. (b) COG.

(0, 0), and R_3 is that defined by the four points $[s_4, IV_1(H)]$, $[1, IV_1(H)]$, (1, 0), and (0.5, 0). Finally, the optimal score value of the fuzzy system (s_5) is calculated from the COG of regions R_1 , R_2 , and R_3 , as shown in Fig. 16(b).

If the output score of the fuzzy system is greater than a threshold, our system determines that user gaze fixation has occurred. Otherwise, our system determines that no gaze fixation has occurred.

3 Experimental Results

Figure 17 shows the experimental setup of our system. In the case that NIR illuminator is set at the left or right position of camera, shadow happens in the opposite side of the eye (compared to the illuminator) because the eye has 3-D spherical shape (not 2-D plane). For example, if the illuminator is set at the left position of the camera, there exists the shadow in the right side of the eye. In this case, the pupil boundary in the shadow region becomes less distinctive, and correct detection of pupil area is difficult. In the case that the NIR illuminator is set at the above position of the camera, the consequent position of the camera becomes lower (compared to the case of our system of Fig. 17) in order not to hide the monitor. In this case, because the camera captures user's eye at a too low position, the vertical resolution of user's eye becomes lower and pupil region is shown more distorted in the vertical direction in the image, which causes the error of pupil detection and measuring change in gaze (Δd) in vertical direction (feature 2). Therefore, the NIR illuminator is positioned below the camera in our gaze tracking system.

We can consider the ring-type illuminator surrounding the camera lens in order to reduce the distance between the camera and NIR illuminator of Fig. 17. However, the phenomenon that the pupil becomes brighter in the captured image (named as "red-eye effect") occurs, which frequently happens in case that the distance between camera and illuminator is too small compared to the distance between camera and user.⁴⁹ If the pupil becomes brighter in the image, correct detection of pupil area is difficult. Therefore, we do not use the ring-type illuminator surrounding the camera lens in our gaze tracking system.

To verify our classification method of gaze fixation and nonfixation, we conducted experiments with 15 participants.

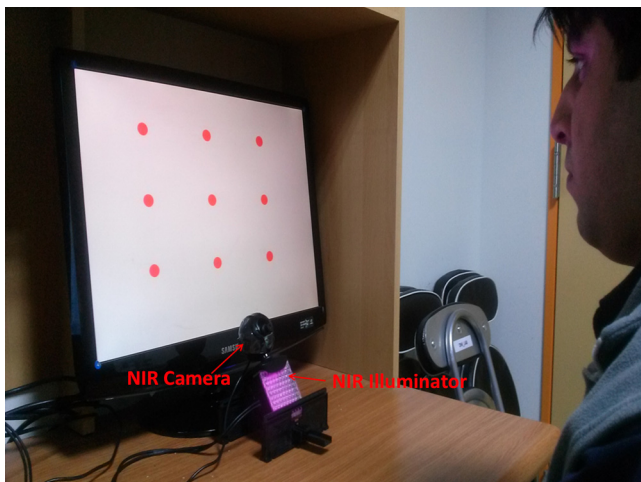


Fig. 17 Experimental setup for the proposed method.

Each person conducted five trials in which they looked at an object of interest in nine positions on a 19-in. monitor, as shown in Fig. 17. The screen resolution is 1680 × 1050 pixels on 19-in. monitor. The size of the circular target is 34 pixels for radius (9 mm for radius). The interdistances between the centers of two circular targets are 453 pixels (120 mm) and 302 pixels (80 mm) in horizontal and vertical directions, respectively, which are the minimum spacing between two objects for our method to distinguish fixation or nonfixation reliably. From this experimental environment, we collected 675 gaze fixation data [true positive (TP) data] and the same number of nonfixation data [true negative (TN) data]. The TP data were collected when each participant looked at the nine positions with the intention of activating or selecting the object. The TN data were collected when each participant looked at positions away from the object of interest with the intention of simply looking at these regions.

To measure the accuracy of the classification of gaze fixation and nonfixation with these TP and TN data, we compared the equal error rate (EER) with different defuzzification methods. We considered type I errors, where TP data were incorrectly classified as TN, and type II errors, where TN data were incorrectly classified as TP. As explained in Sec. 2.5.3, if the output score of the fuzzy system is greater than a threshold, our system determines that user gaze fixation has occurred (TP). Otherwise, our system determines that user gaze fixation has not occurred (TN). Therefore, the type I and II errors change according to the threshold. With a larger threshold, the prevalence of type I errors increases, whereas that of type II errors decreases. Conversely, with

Table 4 Classification results of gaze fixation and nonfixation using MIN rule (unit: %).

Defuzzification method	Threshold	Type I error	Type II error	EER
FOM	0.50	0	0.83	0.42
MOM	0.50	0	0.83	0.42
LOM	0.75	0	0.83	0.42
MeOM	0.50	0	0.83	0.42
COG	0.70	0.17	0	0.09

Table 5 Classification results of gaze fixation and nonfixation using MAX rule (unit: %).

Defuzzification method	Threshold	Type I error	Type II error	EER
FOM	0.50	17.83	0	8.92
MOM	0.50	17.83	0	8.92
LOM	0.92	11	10	10.5
MeOM	0.50	17.83	0	8.92
COG	0.58	0.17	0	0.09

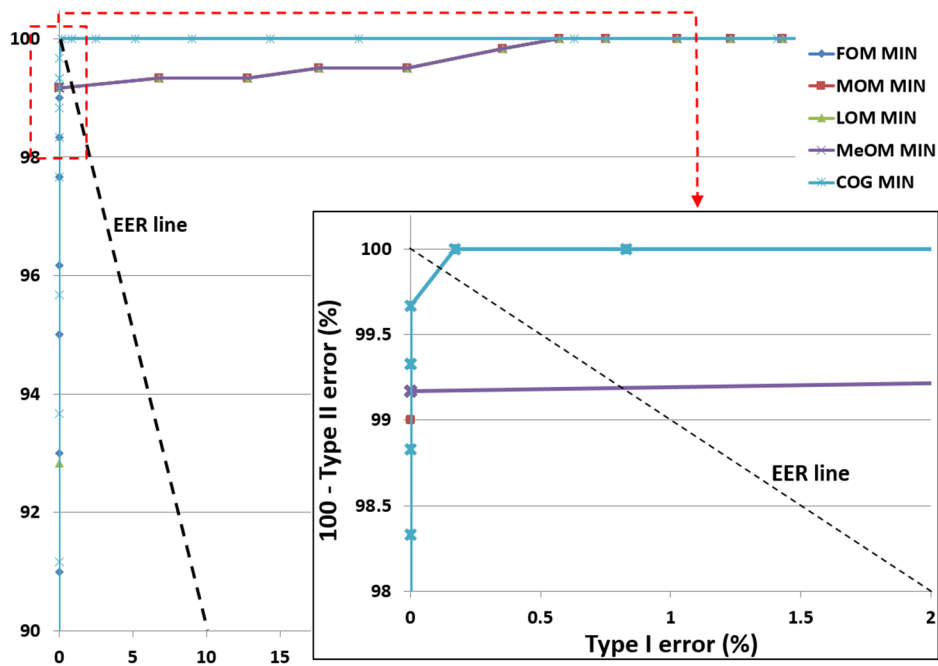


Fig. 18 ROC curves from classification of gaze fixation and nonfixation according to various defuzzification methods with MIN rule.

a smaller threshold, the number of type I errors decreases and the number of type II errors increases. EER is usually calculated by averaging the type I and II errors when the threshold is such that they have a similar prevalence.

The classification results of gaze fixation and nonfixation given by the five defuzzification methods using the MIN and MAX rules are listed in Tables 4 and 5, respectively. As indicated in these tables, the smallest EER (~0.09%) was obtained by COG with both the MIN and MAX rules.

Figures 18 and 19 show the receiver operating characteristic (ROC) curves for the classification results of gaze fixation and nonfixation according to the various defuzzification methods using the MIN and MAX rules, respectively. As shown in these figures, the classification accuracy of COG with the MIN and MAX rules is higher than those achieved by the other defuzzification methods.

In the ROC curves of Figs. 18 and 19, we show the changes of “100—type II error (%)” according to the

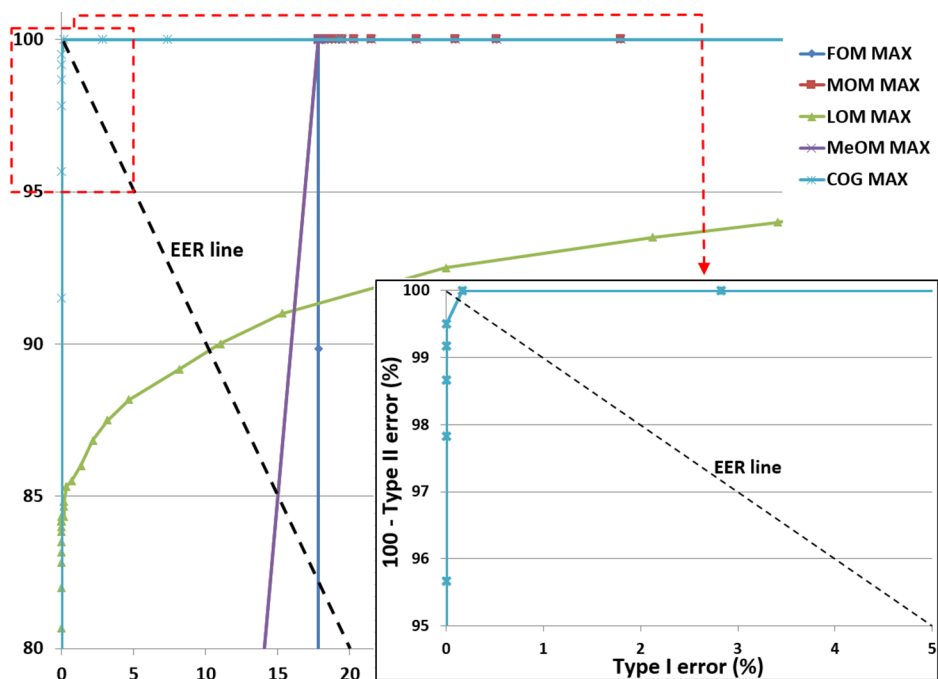


Fig. 19 ROC curves from classification of gaze fixation and non-fixation of gaze according to various defuzzification methods with MAX rule.

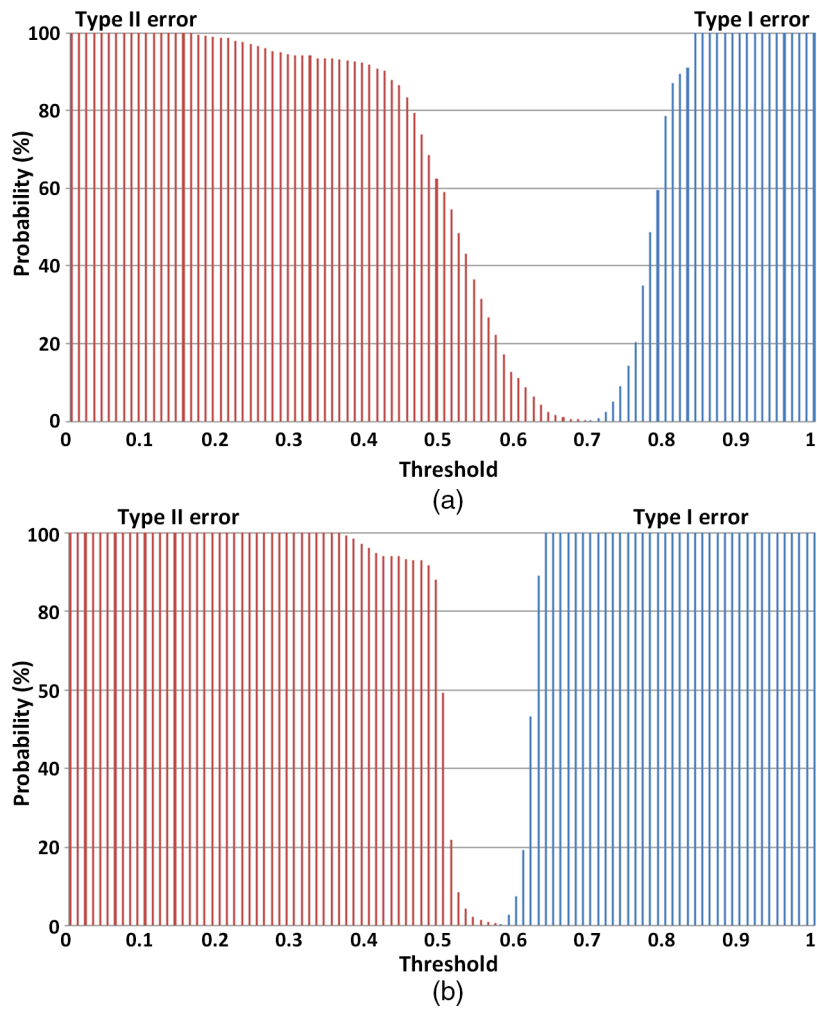


Fig. 20 The type I and II errors according to threshold. (a) COG method with MIN rule. (b) COG method with MAX rule.

increase of type I error (%). The left-upper position of the graphs is (0, 100) (“type I error” of 0% and “100—type II error” of 100%). Because “100—type II error” is 100%, the consequent type II error is 0%. Therefore, the left-upper

position of the graphs represents the position of no error of type I and II. From that, we can know that the ROC curves closer to the left-upper position (COG MIN of Fig. 18 and COG MAX of Fig. 19) than others show the lower errors of type I and II (higher accuracies of classification of gaze fixation and nonfixation). EER is usually calculated by averaging the type I and II errors when the threshold is such that they (type I and II errors) have a similar prevalence. Therefore, EER line is that passing through the two points where types I and II errors are same. For example with Fig. 18, these two points are (0, 100) and (2, 98). Because 100 and 98 represent the “100—type II error”, the consequent type II errors are 0 and 2, respectively. Therefore, two points are (0, 0) and (2, 2) in terms of type I and II errors, respectively.

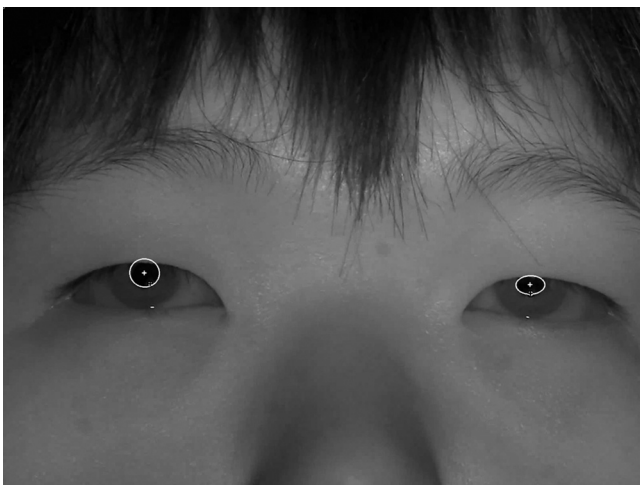


Fig. 21 Example of incorrect detection of pupil boundary and center, which causes type I errors.

Table 6 EER comparison between our method and previous method (unit: %).

Method	EER
Proposed method	0.09
Dwell time-based method	1.53

Table 7 Processing time for our proposed method (unit: ms).

Detecting pupil and glint centers	Calculating pupil size	Calculating feature 1	Calculating feature 2	Fuzzy based determination of gaze fixation	Total
29.831	0.784	1.090	0.005	0.017	31.727

Figure 20 shows the type I and II errors according to threshold using COG with the MIN and MAX rules. It can be seen that the proposed method produces the small common area of type I and II errors, which shows that the EER by our method is low.

As shown in Tables 4 and 5, the proposed method with COG gave type I errors in 0.17% of cases, whereas it produced no type II errors. As explained before, type I errors occur when TP data are incorrectly classified as TN. Type I errors occurred for the following reason. In people whose pupil is partially occluded by their eyelid, an incorrect pupil boundary can be detected (as shown in the right eye of Fig. 21) compared to the left eye, which causes the incorrect pupil center to be detected. In our approach, the final gaze position is calculated by averaging the gaze positions of both eyes. Therefore, incorrect detection of the right pupil center can cause incorrect gaze detection. The pupil center may be correctly detected in one image and incorrectly detected in the next image because of occlusion of the pupil by the eyelid. The consequent gaze position will fluctuate, causing the change in gaze (Δd) of Eq. (7) to increase and resulting in a type I error.

In a second experiment, we compared the performance of our proposed method with that of a popular approach based on the dwell time.¹³ This comparison uses the same 675 gaze fixation data (TP data) and 675 nonfixation data (TN data) obtained from the 15 participants in the first experiment. As indicated in Table 6, our method outperformed the previous method in the classification of gaze fixation and nonfixation.

Furthermore, we analyzed the processing time of our proposed method on a desktop computer with a 2.5-GHz CPU and 4-GB memory. The results are presented in Table 7. The proposed method required a total processing time of ~31 ms, most of which is dedicated to detecting the pupil and glint centers. These results confirm that our method can operate at fast speeds [~ 31.5 fps (= $1000/31.727$)].

In our research, 15 people took part in the experiments, and each person conducted five trials. The ages of the people ranged from 24 to 45. Five people wore contact lens, and it did not affect the experimental results, six female, and nine male. From the experiments, we can confirm that gender did not affect the experimental results, either. Each participant was not requested to take a rest before experiments, and they were randomly selected without the preparation for experiment. Therefore, people having various mental or physical state took part in the experiments, which show that mental or physical state did not give much effect on the results, either. Nevertheless, the pupil accommodation can be affected by the change of environmental lighting conditions⁵⁰ and psychological effect such as severe auditory emotional (negative or positive) stimuli.⁵¹ Therefore, environmental light was not changed by being maintained as about 350 lux. (considering conventional office environment⁵²) and any severe auditory emotional stimuli was not given to people during the experiments because it is not often the case

with the frequent change of environmental light and severe auditory emotional stimuli assuming that our system is used in conventional office environment. However, in general, the speed of pupil size change is reported to be lower with the old people compared to the young people.⁵³ Therefore, in case our system is used for the old people whose age is over 50- or 60-years old, the Peakedness (P_k) of Eq. (2) can be measured with the time window of larger size than the original one [W of Eq. (2)].

4 Conclusions

In this study, we have developed a determination method for gaze fixation in NIR camera-based gaze tracking systems. We employed two features, i.e., the change in pupil size (for measuring pupil accommodation) and change in gaze position over a short dwell time. A fuzzy system was adopted using these two features as input values, and the gaze fixation or nonfixation was determined through defuzzification. The performance of the proposed method was investigated by comparing the defuzzification results with ROC curves and EER. From the results, we verified that the COG method with MIN and MAX rules outperformed other methods in terms of accuracy and that our system can operate at fast speeds.

In future work, we intend to enhance the performance of gaze fixation by combining the features of the change in pupil size and change in gaze position with texture information from the target region.

Acknowledgments

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1D1A1A01056761).

References

1. R. A. Bolt, "Eyes at the interface," in *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, Gaithersburg, Maryland, 15–17 March 1982, pp. 360–362 (1982).
2. M. A. Just and P. A. Carpenter, "The role of eye-fixation research in cognitive psychology," *Behav. Res. Meth. Instrum.* **8**, 139–143 (1976).
3. C. Mauri et al., "Computer vision interaction for people with severe movement restrictions," *Interdiscip. J. Hum. ICT Environ.* **2**, 38–54 (2006).
4. "The device to tell the will. Let's chat," <http://panasonic.biz/healthcare/aflt/products/letschat/#anc-02> (12 November 2015).
5. C. G. Pinheiro, Jr. et al., "Alternative communication systems for people with severe motor disabilities: a survey," *Biomed. Eng. Online* **10**, 31 (2011).
6. B. Rebsamen et al., "Controlling a wheelchair indoors using thought," *IEEE Intell. Syst.* **22**, 18–24 (2007).
7. C. Choi and J. Kim, "A real-time EMG-based assistive computer interface for the upper limb disabled," in *Proc. of the IEEE 10th Int. Conf. on Rehabilitation Robotics*, Noordwijk, The Netherlands, 12–15 June 2007, pp. 459–462 (2007).
8. L. Y. Deng et al., "EOG-based human-computer interface system development," *Expert Syst. Appl.* **37**, 3337–3343 (2010).
9. R. Barea et al., "EOG-based eye movements codification for human computer interaction," *Expert Syst. Appl.* **39**, 2677–2683 (2012).
10. C.-S. Lin et al., "Powered wheelchair controlled by eye-tracking system," *Opt. Appl.* **36**, 401–412 (2006).

11. T. Kocejko, A. Bujnowski, and J. Wtorek, "Eye mouse for disabled," in *Proc. of Conf. on Human System Interactions*, Krakow, Poland, 25–27 May 2008, pp. 199–202 (2008).
12. M.-C. Su, K.-C. Wang, and G.-D. Chen, "An eye tracking system and its application in aids for people with severe disabilities," *Biomed. Eng. Appl. Basis Commun.* **18**, 319–327 (2006).
13. H. Heo et al., "Nonwearable gaze tracking system for controlling home appliances," *Sci. World J.* **2014**, 1–20 (2014).
14. B. Ashtiani and I. S. MacKenzie, "BlinkWrite2: an improved text entry method using eye blinks," in *Proc. of the Symp. on Eye-Tracking Research and Applications*, Austin, Texas, 22–24 March 2010, pp. 339–346 (2010).
15. M. J. Doughty, "Further assessment of gender- and blink pattern-related differences in the spontaneous eyeblink activity in primary gaze in young adult humans," *Optom. Vision Sci.* **79**, 439–447 (2002).
16. R. J. K. Jacob, "Eye movement-based human-computer interaction techniques: toward non-command interfaces," in *Advances in Human-Computer Interaction*, Ablex Publishing Co., Norwood, New Jersey (1993).
17. J. P. Hansen et al., "Gaze typing compared with input by head and hand," in *Proc. of the Symp. on Eye Tracking Research and Applications*, San Antonio, Texas, 22–24 March 2004, pp. 131–138 (2004).
18. A. J. Hornof and A. Cavender, "EyeDraw: enabling children with severe motor impairments to draw with their eyes," in *Proc. of the Conf. on Human Factors in Computing Systems*, Portland, Oregon, 2–7 April 2005, pp. 161–170 (2005).
19. A. Huckauf and M. H. Urbina, "Object selection in gaze controlled systems: what you don't look at is what you get," *ACM Trans. Appl. Percept.* **8**, 13 (2011).
20. Á. Kristjánsson, M. W. G. Vandenberghe, and J. Driver, "When pros become cons for anti-versus prosaccades: factors with opposite or common effects on different saccade types," *Exp. Brain Res.* **155**, 231–244 (2004).
21. C. Ware and H. H. Mikaelian, "An evaluation of an eye tracker as a device for computer input," in *Proc. of the SIGCHI/GI Conf. on Human Factors in Computing Systems and Graphics Interface*, Toronto, Ontario, Canada, 5–9 April 1987, pp. 183–188 (1987).
22. S. Zhai, C. Morimoto, and S. Ihde, "Manual and gaze input cascaded (MAGIC) pointing," in *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, Pittsburgh, Pennsylvania, 15–20 May 1999, pp. 246–253 (1999).
23. M. Kumar, A. Paepcke, and T. Winograd, "EyePoint: practical pointing and selection using gaze and keyboard," in *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, San Jose, California, 28 April–3 May 2007, pp. 421–430 (2007).
24. K. Grauman et al., "Communication via eye blinks and eyebrow raises: video-based human-computer interfaces," *Univ. Access Inf. Soc.* **2**, 359–373 (2003).
25. M. Kaur et al., "Where is 'it'? Event synchronization in gaze-speech input systems," in *Proc. of the Fifth Int. Conf. on Multimodal Interfaces*, Vancouver, British Columbia, Canada, 5–7 November 2003, pp. 151–158 (2003).
26. V. Surakka, M. Illi, and P. Isokoski, "Gazing and frowning as a new human-computer interaction technique," *ACM Trans. Appl. Percept.* **1**, 40–56 (2004).
27. O. Tuisku et al., "Text entry by gazing and smiling," *Adv. Hum. Comput. Interact.* **2013**, 1–13 (2013).
28. R. J. K. Jacob, "The use of eye movements in human-computer interaction techniques: what you look at is what you get," *ACM Trans. Inf. Syst.* **9**, 152–169 (1991).
29. S. P. Verney, E. Granholm, and S. P. Marshall, "Pupillary responses on the visual backward masking task reflect general cognitive ability," *Int. J. Psychophysiol.* **52**, 23–36 (2004).
30. E. H. Hess, *The Tell-Tale Eye: How Your Eyes Reveal Hidden Thoughts and Emotions*, 1st ed., Van Nostrand Reinhold Co., New York (1975).
31. "Webcam C600," <http://www.logitech.com/en-us/support/webcams/5869> (17 November 2015).
32. Y. He, "Key techniques and methods for imaging iris in focus," in *Proc. of the IEEE Int. Conf. on Pattern Recognition*, Hong Kong, China, 20–24 August 2006, pp. 557–561 (2006).
33. "LC Technologies, Inc.," <http://www.eyegaze.com/> (29 March 2016).
34. "SMI," <http://www.smivision.com/> (29 March 2016).
35. "TheEyeTribe," <https://theyetribe.com/> (29 March 2016).
36. "Tobii," <http://www.tobii.com/> (29 March 2016).
37. B.-S. Kim, H. Lee, and W.-Y. Kim, "Rapid eye detection method for non-glasses type 3D display on portable devices," *IEEE Trans. Consum. Electron.* **56**, 2498–2505 (2010).
38. P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vision* **57**, 137–154 (2004).
39. R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed., Prentice-Hall, New Jersey (2002).
40. L. Ding and A. Goshtasby, "On the canny edge detector," *Pattern Recogn.* **34**, 721–725 (2001).
41. F. Edward Su, "Area of an ellipse," <https://www.math.hmc.edu/funfacts/ffiles/10006.3.shtml> (17 November 2015).
42. "Moving average," https://en.wikipedia.org/wiki/Moving_average (17 November 2015).
43. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge (2000).
44. B. S. Bayu and J. Miura, "Fuzzy-based illumination normalization for face recognition," in *Proc. of the IEEE Workshop on Advanced Robotics and Its Social Impacts*, Tokyo, Japan, 7–9 November 2013, pp. 131–136 (2013).
45. A. Barua, L. S. Mudunuri, and O. Kosheleva, "Why trapezoidal and triangular membership functions work so well: towards a theoretical explanation," *J. Uncertain Syst.* **8**, 164–168 (2014).
46. J. Zhao and B. K. Bose, "Evaluation of membership functions for fuzzy logic controlled induction motor drive," in *Proc. of the IEEE Annual Conf. of the Industrial Electronics Society*, Sevilla, Spain, 5–8 November 2002, pp. 229–234 (2002).
47. W. V. Leekwijck and E. E. Kerre, "Defuzzification: criteria and classification," *Fuzzy Sets Syst.* **108**, 159–178 (1999).
48. E. V. Broekhoven and B. D. Baets, "Fast and accurate center of gravity defuzzification of fuzzy system outputs defined on trapezoidal fuzzy partitions," *Fuzzy Sets Syst.* **157**, 904–918 (2006).
49. "Red-eye effect," https://en.wikipedia.org/wiki/Red-eye_effect (26 May 2016).
50. P. Reeves, "The response of the average pupil to various intensities of light," *J. Opt. Soc. Am.* **4**, 35–43 (1920).
51. T. Partala and V. Surakka, "Pupil size variation as an indication of affective processing," *Int. J. Hum. Comput. Stud.* **59**, 185–198 (2003).
52. Lux, <https://en.wikipedia.org/wiki/Lux> (26 May 2016).
53. R. Feinberg and E. Podolak, *Latency of Pupillary Reflex to Light Stimulation and Its Relationship to Aging*, pp. 1–14, Federal Aviation Agency, Washington D.C. (1965).

Rizwan Ali Naqvi received his BS degree in computer engineering from Comsats Institute of Technology, Pakistan, in 2008. He received his MS degree in electrical engineering from Karlstad University, Sweden, in 2011. He was a lecturer at Superior University, Pakistan, until 2014. Currently, he is a PhD candidate in the Department of Electronics and Electrical Engineering at Dongguk University. His research interests include image processing, computer vision, and HCI.

Kang Ryoung Park received his BS and MS degrees in electronic engineering from Yonsei University, Seoul, Korea, in 1994 and 1996, respectively. He received his PhD in electrical and computer engineering from Yonsei University in 2000. He has been a professor in the Division of Electronics and Electrical Engineering at Dongguk University since March 2013. His research interests include computer vision, image processing, and biometrics.