Article

Fuzzy System-Based Target Selection for a NIR Camera-Based Gaze Tracker

Rizwan Ali Naqvi, Muhammad Arsalan and Kang Ryoung Park *

Division of Electronics and Electrical Engineering, Dongguk University, 30 Pildong-ro 1-gil, Jung-gu,
Seoul 100-715, Korea; rizwanali@dongguk.edu (R.A.N.); arsal@dongguk.edu (M.A.)

* Correspondence: parkgr@dongguk.edu; Tel.: +82-10-3111-7022; Fax: +82-2-2277-8735

Abstract: Gaze-based interaction (GBI) techniques have been a popular subject of research in the last few decades. Among other applications, GBI can be used by persons with disabilities to perform everyday tasks, as a game interface, and can play a pivotal role in the human computer interface (HCI) field. While gaze tracking systems have shown high accuracy in GBI, detecting a user's gaze for target selection is a challenging problem that needs to be considered while using a gaze detection system. Past research has used the blinking of the eyes for this purpose as well as dwell time-based methods, but these techniques are either inconvenient for the user or requires a long time for target selection. Therefore, in this paper, we propose a method for fuzzy system-based target selection for near-infrared (NIR) camera-based gaze trackers. The results of experiments performed in addition to tests of the usability and on-screen keyboard use of the proposed method show that it is better than previous methods.

1. Introduction

The field of gaze-based interaction (GBI) has witnessed a significant growth in recent years, in response to certain long-standing challenges in gaze-tracking research. GBI helps persons with disabilities communicate with other people or devices. In 1982, Bolt first showed that the gaze can facilitate human-computer interface (HCI) implementation [1]. Using gaze input to perform tasks or computer operations eventually became popular among users with motor disabilities. Mauri et al. proposed the idea of using computer-assistive technologies, i.e., joysticks, trackballs, virtual keyboards, and eye tracking devices, to interact with computers [2]. These technologies, however, do not work well for severely disabled people who cannot use certain parts of their body. Alternative communication devices such as electromyogram (EMG), electroencephalogram (EEG), and electro-oculogram (EOG) [3,4] are better options, but are expensive and unavailable to most people. They may also be frustrating for users as they require electrodes to be attached to the body. Hence, camera-based gaze detection methods are a desirable alternative. Gaze tracking methods on the 2D monitor of a desktop computer have been widely studied [5–7]. However, there are some limitations to these methods, i.e., the inability to control 3D space, and significant degradation in accuracy due to variations in the Z-distance between the monitor and a user's eyes. Therefore, to control home appliances in 3D space, a non-wearable gaze tracking system has been proposed [8]. Increasing the accuracy of gaze detection has primarily been emphasized in past studies on gaze detection, but few have tackled target selection using a detection system. To the best of our knowledge, past research used methods based on the clicks of a mouse or a keyboard, the dwell time of the gaze position, and target selection by number of eye blinks. However, these methods have limitations in terms of selection speed as well as user convenience. A detailed summary past research on target selection in gaze trackers is provided in the next section.

2. Related Work

To select the target of interest according to a user's intention, approaches to locating the target of his/her gaze can be categorized into gaze-based and visual saliency-based methods. The former category includes eye blinks, dwell time, on- and off-screen buttons, keystrokes, eyebrow raises, and speech. Blinking can be used to select letters from the alphabet that can solve issues in eye typing where gaze direction is utilized to point to these letters [9]. Blinking for commands and normal blinking are usually difficult to discriminate [10]; thus, eyes that are closed for a lengthy period may be required for such discrimination, which can affect performance and the user's convenience. The dwell time-based method seems to be a better and more natural approach compared to the blinking-based method [11]. However, for this approach, the gaze tracking system requires knowledge of the position and duration of the user's gaze on the selected object. Past research [12–14] has considered methods that use the dwell time of the position of a user's gaze. Further, an approach proposed by Huckauf et al., instead of using dwell time and blink-based selection, accomplishes target selection by using antisaccades [15]. Ware et al. proposed selecting the object of interest by fixation and subsequent saccade towards on/off buttons [16].

In past research [11–23], one modality has been used to point to or to select the object of interest. However, these methods typically suffer from problems whereby the object is selected every time the user looks at it, intentionally or unintentionally. This problem was first pointed out as the "Midas touch problem" [17]. It needs to be avoided, and the object of interest to a user should be discriminated from unintentionally gazes at points. The selection of the object of interest by eye blink encounters difficulty in discriminating intentional from unintentional blinks. Likewise, dwell time-based selection encounters the same type of problem, i.e., if the dwell time is very short, it faces the same problem described above: the Midas touch problem; on the other hand, if dwell time is too long, it can degrade performance as well as become tire the user [18]. A possible solution is to use graphical on/off screen buttons, but this too can be problematic, as it can interfere with the user's intention with regard to the relevant object and deviate from the object of interest. Some past work [19,20] has proposed the use of manual inputs, such as keystrokes, combined with gaze control for pointing at and selecting objects of interest. Surakka et al. introduced the idea of a frowning face to selecting the object of interest [21]. The method based on eye blinks or eyebrow raises was proposed by Grauman et al., where these were used to point at and select the object and convey the relevant command [22]. Tuisku et al. used gazing and smiling to this end, where gazing was used to point to the object and smiling as a selection tool [23]. Although these techniques performed well for their intended purposes, they have limited accuracy and speed of selection according to the user's intention.

To solve the problems of single modality-based methods, multiple modality-based methods need to be explored. The authors of [24] proposed the idea of a multiple modality-based method based on pupil accommodation and dwell time. However, because they did not consider information from the monitor image (textural or high-frequency information in the area of the monitor image at which a user gazes), there is room for further improvement in detecting the user's gaze for target selection. Furthermore, the accurate measurement of pupil accommodation and dwell time is dependent on the correct detection of pupil size and the centers of the pupil and the corneal glint. Therefore, incorrect detection can affect system performance.

Researchers have also shown considerable interest in visual attention models that can be implemented to classify of intentional and unintentional gazes [25]. Saliency-based models have been used to measure the possibility of a location to attract the observer's attention. A saliency map can be obtained from a visual image for gaze points. Attaining visual information is a low-cost preprocessing step through which visual systems (biological or artificial) select the most noticeable information from a given scene [26]. Saliency prediction with higher accuracy has a number of applications in salient object detection, automatic vehicle guidance, scene understanding, robot navigation, specifically fast regions of interest (ROI) selection in complex visual scenes, and so on [27]. A large number of saliency-based methods for detecting intentional gazes have been proposed [28–36]. Most models of

saliency are based on bottom-up image cues as they are biologically inspired. Many theoretical models are based on the feature integration theory proposed by Treisman and Gelade [37]. They analyzed the visual features that are combined to direct human attention over conjunction search tasks. A feed forward model to combine these features with the concept of saliency map was proposed by Koch and Ullman [38], which was implemented and verified by Itti et al. [39]. Different saliency models are based on low-, middle-, and high-level features. All models have different assumptions and methodologies focusing on different aspects of human visual behavior. These models can mainly be categorized into bottom-up, top-down, and learning-based approaches [27].

In the bottom-up approach [39–45], saliency models use biologically reasonable low-level features depend on computational principles proposed by Itti et al. [39]. On the basis of image features, i.e., color, intensity, and orientation, they derived bottom-up visual saliency based on center-surround differences across multi-scale image features. Bruce and Tsotsos [40] proposed attention for an information maximization (AIM) saliency model, but it is weak in detecting local dissimilarities (i.e., local vs. global saliency). A graph-based visual saliency (GBVS) model was proposed by Harel et al. based on the assumption that image patches that are different from surrounding patches are salient [41]. Like Itti et al.'s method, GBVS also fails to detect global visual saliency. Liu et al. [42] identified ROI by conditional random fields (CRF) using three types of features: multi-scale contrast, center-surround histogram, and color spatial distribution. Kong et al. proposed the idea of a multi-scale integration strategy to combine various low-level saliency features [43]. Fang et al. proposed compressed domain saliency detection using color and motion characteristics [44]. Jiang et al assumed that the area of interest of visual attention has a shape prior, and detected salient regions using contour energy computation [45]. Low-level feature-based methods perform well in general but do not consider semantic information, such as faces, humans, animals, objects, and text.

To solve this problem, the top-down approach has been researched [46–49]. Several studies have adopted this approach and attained performance improvement by adding high-level features. Cerf et al. [46] added a high-level factor, face detection, to Itti's model and showed that it enhances performance. Judd et al. [47] proposed a saliency detection model by learning the best weights for all combined features using support vector machines (SVMs). Chang et al. [48] subsequently proposed the idea of an object-based saliency model by relying on the claim that observers tend to look at the centers of objects. Hence, top-down models have highlighted the importance of high-level and semantic features, e.g., faces, animals, the image-centric bias, and the object-centric bias. A major problem with this approach is that it is training dependent, and often fails to detect salient features if its models are not trained. To consider the limitations of the bottom-up and top-down approaches, the learning-based method has been researched [50–57]. An AdaBoost-based model for feature selection was proposed by Borji [52]. It models complex input data by combining a series of base classifiers. Reingal et al. observed that there were different statistics for fixated patches compared to random patches [53]. Borji et al. [56] later combined two simple methods (multiplication and sum) using normalization schemes (identity, exponential, and logarithmic) to combine saliency models. The evolutionary optimization algorithm has been used by some researchers to find optimal sets of combination weights [57].

All saliency-based methods can provide only information concerning regions where people tend to gaze with higher probability, instead of accurate intentional gaze position. Therefore, to overcome the limitations of past work on gaze-based methods as well as visual saliency-based methods, we propose a fuzzy system-based target selection method for near-infrared (NIR) camera-based gaze trackers by fusing the gaze-based method with the bottom-up visual saliency-based method. Our method combines multi-modal inputs, i.e., pupil accommodation measured by template matching, short dwell time, and Gabor filtering-based texture information of visual saliency using a fuzzy system. In the following four ways, our research is novel compared to past research. The first, second, and third points represent major novelties whereas the fourth is a minor one:

- First, a new and improved method based on the Chan–Vese algorithm is proposed for detecting the pupil center with a boundary as well as the glint center.
- Second, we use three features, i.e., change in pupil size using template matching (to measure pupil accommodation), change in gaze position during short dwell time, and Gabor filtering-based texture information of monitor image at gaze target. A fuzzy system is then used with these three features as inputs, and the decision concerning the user's target selection is taken through defuzzification.
- Third, an optimal input membership function for fuzzy system can be obtained based on the maximum entropy criterion.
- Fourth, through comparative experiments using an on-screen keyboard on a previous dwell time-based method, the performance and usability of our method were verified in a real gaze-tracking environment.

In Table 1, we summarize the comparison of the proposed with existing methods.

Table 1. Comparison of previous and proposed methods.

Category	Method	Advantages	Disadvantages
Gaze-based methods	Use a single modality, i.e., eye blink [9,22], dwell time [12–14], antisaccades [15], on and off screen buttons [16], key strokes [19,20], face frowning [21], eye brow raises [22], and smiling [23]	- Single sensing modality is used, and so is simpler in design compared to the method based on multiple modalities - Fewer input data items are required compared to multiple modalities	- Lower accuracy of selection of target than methods based on multiple modalities - Completely dependent on a single modality; minor errors in sensing data can badly affect overall results - Some methods using a single modality are not feasible for users with high levels of motor disabilities, especially for those who can only move their eyes [19–23]
	Use multiple modalities, i.e., pupil accommodation and dwell time [24]	- Accuracy of intentional object selection is high compared to methods based on single modality	- There is room for further enhancement of detecting user's gaze for target selection - Incorrect detection of pupil size and the centers of pupil and corneal glint can significantly affect accuracy
Visual saliency-based methods	Bottom-up approach [39–45]	- Performs well for detecting low-level features	- Semantic information is not considered
	Top-down approach [46–49]	- Performs well for high-level features	- Performance can be affected by the correct detection of high-level features, i.e., face detection
	Learning-based method [50–57]	- Efficient for complex input data	- Performs well for particular category of data on which it is trained
Fusing gaze-based method with visual saliency-based method	Combining pupil accommodation with template matching, short dwell time, and the texture information of visual saliency by fuzzy system (**proposed method**)	- Higher accuracy of target selection - Less affected by the incorrect detection of pupil and corneal glint	- More input data needs to be processed compared to gaze-based and visual saliency-based methods

The remainder of this paper is organized as follows: in Section 3, our proposed system and methodology are introduced. The experimental setup is explained and the results are presented in Section 4. Section 5 contains our conclusions and discussion of some ideas for future work.

3. Target Selection by Using Pupil Accommodation with Template Matching, Change in Gaze, and Texture Information from a Monitor Image of Gaze Target

3.1. Overview of Proposed Method

In the proposed method, images of the eye were taken using an image acquisition system on a commercial Web camera, i.e., a Logitech (Lausanne, Switzerland) C600 [58] with a universal serial bus (USB) interface, and near-infrared (NIR) was used as source of illumination, with 8×8 NIR light-emitting diodes (LEDs) for our gaze tracker. The use of NIR illuminator served three important purposes [59,60]: first, it minimized the impact of different ambient light conditions; second, it distinguished the boundary of the pupil; third, as NIR is barely visible, it can minimize any interference while using the device in applications. In detail, the NIR light of shorter wavelength (shorter than 800 nm) has the tendency of making the iris darker (compared to the case using NIR light of wavelength greater than 800 nm). Therefore, the boundary between the pupil and the iris in the image becomes less distinct, and the error associated with locating the pupil boundary increases. On the other hand, NIR light of longer wavelength (longer than 800 nm) has the opposite tendency of making the iris brighter. Therefore, the boundary between the pupil and the iris in the image becomes more distinct, and the error associated with locating the pupil boundary is reduced. However, the camera image sensor is usually less sensitive to light according to the increase in wavelength, which means that the image captured by light of wavelength longer than 900 nm becomes darker, and the correct detection of the pupil boundary is consequently difficult. By considering all these factors, an NIR illuminator of 850 nm was adopted in our gaze tracking system.

Images of 1600×1200 pixels were captured at a rate of 30 frames per second (fps). Larger eye images were required to analyze pupil size variations during the user's gaze for target selection. Thus, we used a zoom lens to obtain these images. For pupil accommodation, we need a gaze tracking system that can measure changes in pupil size. Unfortunately, not all commercial gaze tracking systems provide this function [61–64]. Therefore, we implemented our own system.

Figure 1 shows the flowchart of our proposed system. An image of the user's face is first captured by our gaze tracking camera, and the search area of the left or right eye is defined (as shown in Figure 2b). Within this area, the pupil center with boundary and the glint center are found (see the details in Section 3.2). The bright spot on the corneal surface caused by the NIR light is referred to as glint. In the initial step, the user is instructed to observe four positions on the monitor for user-dependent calibration. Pupil size is then measured based on the accurate boundary of the pupil region (see the details in Section 3.3). The features are then calculated to detect the user's gaze for target selection. Feature 1 (F_1) represents pupil accommodation, i.e., it is measured by template matching with the graph of change in pupil size with respect to time (see the details in Section 3.3). The change in gaze position calculated over a short dwell time is referred to as feature 2 (F_2) (see the details in Section 3.4). Furthermore, the texture information of the image on the monitor at the gaze target is measured by Gabor filters and is used as feature 3 (F_3) (see the details in Section 3.5). These three feature values are combined using a fuzzy system, and the user's gaze for target selection can be detected based on the fuzzy output (see the details in Section 3.6).
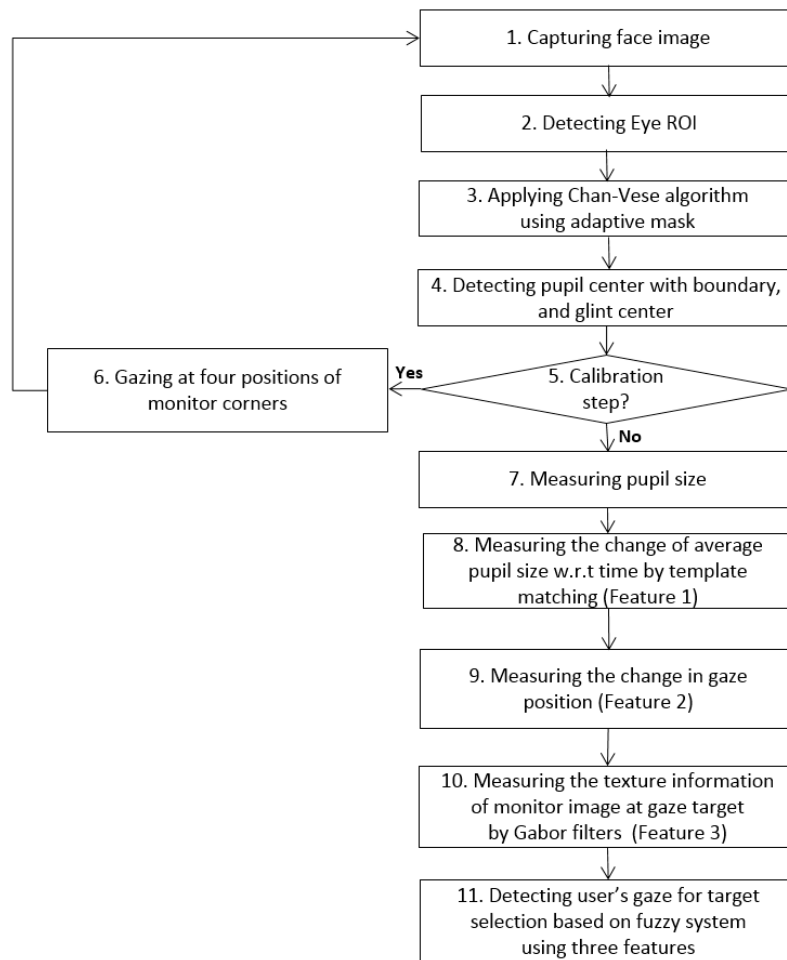
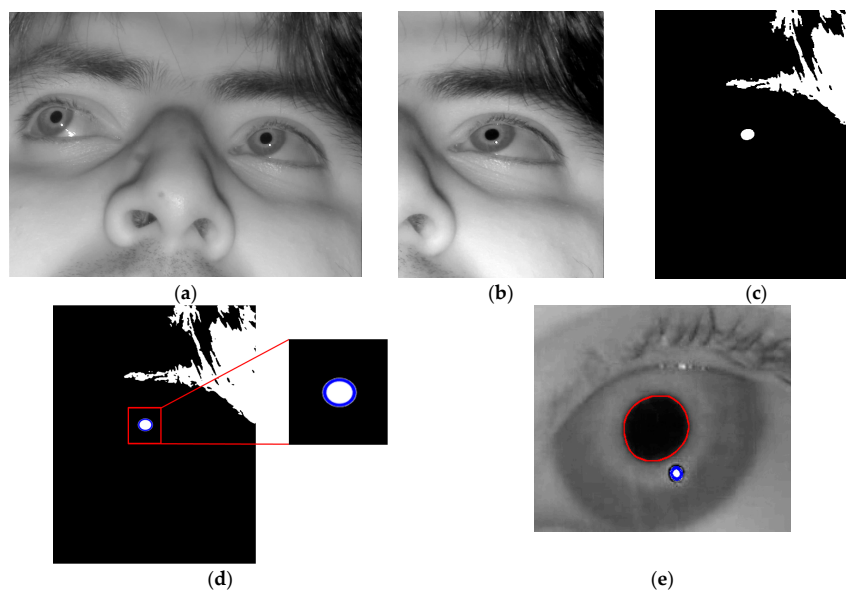**Figure 1.** Overall procedure of the proposed method.

Figure 2. Example of detecting pupil and glint of right eye. (**a**) Input face image; (**b**) Search area of right eye; (**c**) Binarized image; (**d**) Approximate pupil area detected by CHT, and eye ROI defined on the basis of approximate pupil area; (**e**) Pupil and glint boundaries detected by Chan–Vese algorithm with adaptive mask.

3.2. Detection of Pupil and Glint Centers

The correct detection of the pupil region and the glint center is a prerequisite for obtaining accurate feature values for detecting the user's gaze for target selection. In our research, the pupil and glint regions (with geometric centers) were detected as shown in Figure 3, which corresponds to Steps 2–4 of Figure 1.

The approximate pupil area is first extracted as follows: to find dark pixels in the acquired image, binarization based on histogram thresholding is performed (Step 1 in Figure 3). Morphological operations and media filtering operations are executed to remove noise (Step 2 in Figure 3), and the approximate pupil area is detected (Step 3 of Figure 3) based on the circular Hough transform (CHT) [65]. The CHT is a basic technique for detecting circular objects in an image. The concepts of voting and local maxima, i.e., an accumulator matrix, are used to select candidate circles. The ROI including the eye is defined in the input image on the basis of the approximated pupil region (Step 4 of Figure 3). The accurate boundaries of the pupil and the glint are then located within the ROI using the Chan–Vese algorithm [66] based on an adaptive mask (Steps 5–10 of Figure 3). It depends on global properties, i.e., gray-level intensities, contour lengths, and regional areas, rather than local properties such as gradients. The main idea behind the Chan–Vese algorithm is active contour models to evolve a curve from a given image u_o to detect objects in the image. In the classical active contour and snake models [67], an edge detector is used that depends on the gradient of image u_o to stop the evolving curve on the boundary of the required object. By contrast, in the Chan–Vese algorithm, the model is based on trying to separate the image into regions based on intensities. We minimize the fitting terms and add some regularizing terms like the length of the curve C and/or the area of the region inside C. We want to minimize the energy function using level set $\phi(x, y)$ formulation [66]:

$$
\begin{aligned}
C = \partial\omega &= \left\{ (x, y) \, \epsilon \, \Omega : \ \phi(x, y) = 0 \right\} \\
c_{in} = \omega &= \left\{ (x, y) \, \epsilon \, \Omega : \ \phi(x, y) > 0 \right\} \\
c_{out} = \Omega / \overline{\omega} &= \left\{ (x, y) \, \epsilon \, \Omega : \ \phi(x, y < 0) \right\}
\end{aligned}
\tag{1}
$$

where $C = \partial\omega$ is the curve where $\omega \, \epsilon \, \Omega$, and Ω is the planar domain. $c_{in} = \omega$, $c_{out} = \Omega / \overline{\omega}$ represent regions inside and outside curve C, respectively. The energy function $F(c_{in}, \ c_{out}, \ C)$, is defined by:

$$
\begin{aligned}
F(c_{in}, \ c_{out}, \ C) = \mu \cdot Length(C) &+ \ \lambda_{in} \int_{in(C)} |u_o(x, y) - c_{in}|^2 dxdy \\
&+ \lambda_{out} \int_{out(C)} |u_o(x, y) - c_{out}|^2 dxdy
\end{aligned}
\tag{2}
$$

In the above, λ_{in}, λ_{out}, and μ are the positive constants, and $u_o(x, y)$ is the given input image. In order to enhance processing speed and accuracy, we propose an enhanced Chan–Vese algorithm with an adaptive mask based on the approximate pupil region (Steps 9 and 10 of Figure 3) (we use the constraint where (x, y) of Equation (2) belongs to the approximated pupil area). The mask parameters change according to pupil size and location. The CHT provides the rough radius and center of the pupil area of Step 3 in Figure 3. Based on this information, an accurate pupil boundary can be obtained by the Chan–Vese algorithm with adaptive mask, as shown in Steps 10 and 11 of Figure 3. Moreover, the rough size and location of the glint are detected by histogram thresholding and CHT, as shown in Steps 5 and 6 of Figure 3. Based on this information, an accurate glint boundary can be obtained by the Chan–Vese algorithm with adaptive mask as shown in Steps 8 and 11 of Figure 3 (we use the constraint whereby (x, y) of Equation (2) belongs to the approximated glint area). Using the boundaries of the pupil and the glint, their geometric centers are determined as shown in Step 12 of Figure 3.
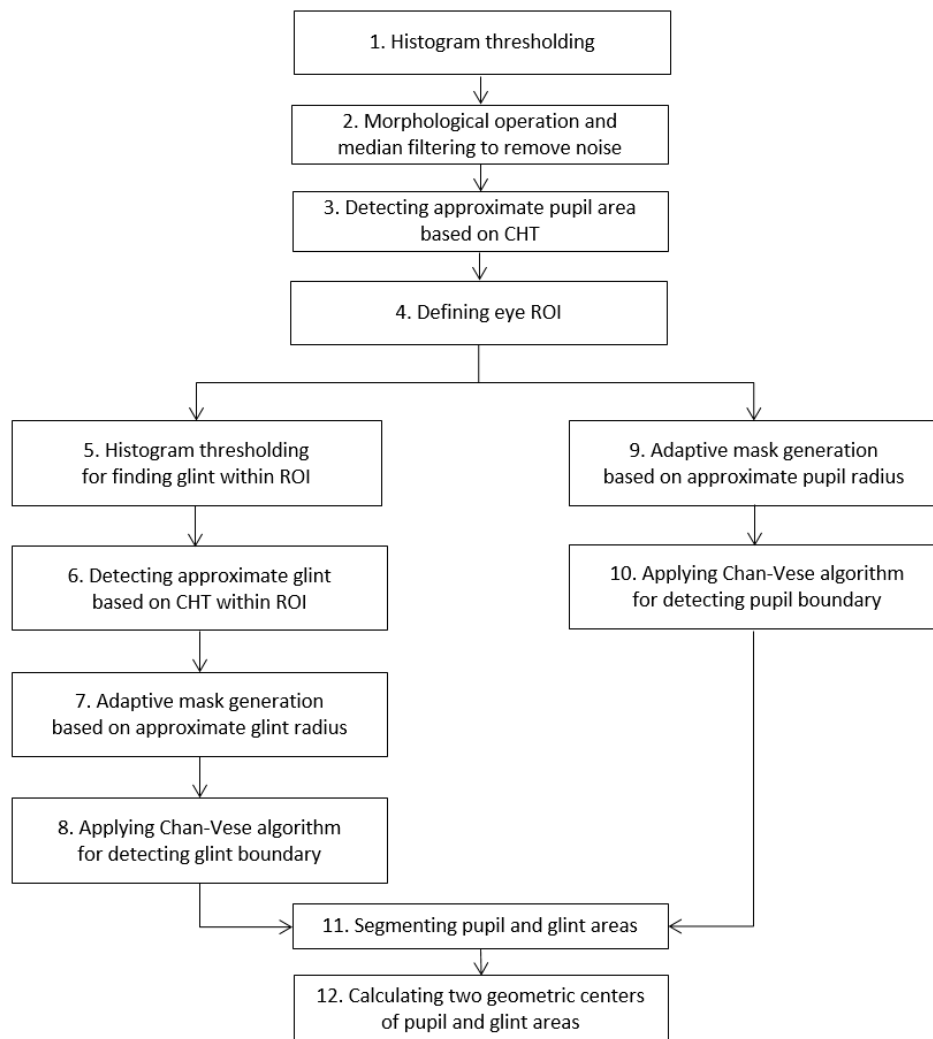
Figure 3. Flowchart for detecting glint center and pupil region.

Figure 2 shows the resulting images of the procedure. Based on these detection results of the pupil and glint areas, two features for inputs to the fuzzy system are calculated, as shown in Sections 3.3 and 3.4.

3.3. Calculating Feature 1 (Change in Pupil Size w.r.t. Time)

As the first feature used to detect the user's gaze for target selection, the change in pupil size with respect to time is measured by template matching. The authors of [24] measured pupil size based on the lengths of the major and minor axes of an ellipse fitted along the pupil boundary. The actual shape of the pupil is not elliptical, as shown in Figure 2e, and, according to [68], pupil detection based on ellipse fitting can yield incorrect pupil size. Therefore, in our research, pupil size is measured by counting total number of pixels inside the pupil boundary detected by the enhanced Chan–Vese method. A graph representing the in pupil size with respect to time is then constructed, and a moving average filter consisting of three coefficients is applied to it to reduce noise.

As mentioned above, the speed of image acquisition of our gaze tracking camera was 30 frames per second. Hence, an image frame was captured in 33.3 ms (1/30 s). To increase the speed of target selection, we used a window of size 10 frames (approximately 333 ms) to measure pupil dilation and constriction over time. Even with a short time window, our method can detect the user's gaze for target selection because three features are simultaneously used.

Past research has shown that cognitive tasks can affect changes in pupil size [69,70]. Based on this concept, the size of the pupil usually decreases in case the gaze is used for target selection, and we can define the shapes resulting from changes in pupil size as shown in Figure 4. Using these shapes, template-based matching is performed with a given window size. Template-based matching is the measure of the similarity between the graph of dilation and constriction in pupil size with respect to time and the expected graph of pupil behavior (template graph shown in Figure 4) employing the user's gaze for target selection. Template-based matching compares these two graphs using the sum of square differences (SSD). The SSD and the final template matching score can be obtained by the following equations:

$$\text{SSD}(p_j, q_{ij}) = \sum_{j=1}^{10} \left(p_j - q_{ij} \right)^2 \tag{3}$$

$$\text{Template matching score} = \min_{i=1,\,2,\,3} \text{SSD}\left(p_j, q_{ij} \right) \tag{4}$$

In the above, p_j is the jth value of pupil size in the graph of the input, and q_{ij} is the jth value of the pupil size of the ith template graph. We used three template graphs as shown in Figure 4. The starting position of the template matching window was detected based on changes in gaze position, i.e., change in the horizontal and vertical gaze directions. The template matching score of Equation (1) decreased in case the user's gaze was employed for target selection and increased in other cases. The template matching score was used as feature 1. Since the maximum and minimum values of pupil size can show individual variation according to people, the value of pupil size was normalized by min-max scaling before template matching, as shown in Figure 4. Although one previous study [24] used a similar concept to measure the change in pupil size through peakedness, this method has the disadvantage whereby the accurate position of the peak needs to be detected in advance in the graph for pupil size, which can be affected by noise in the input data. In contrast to this, our method does not need to detect the position of the peak, and is more robust against noise in the input data (see Section 4).
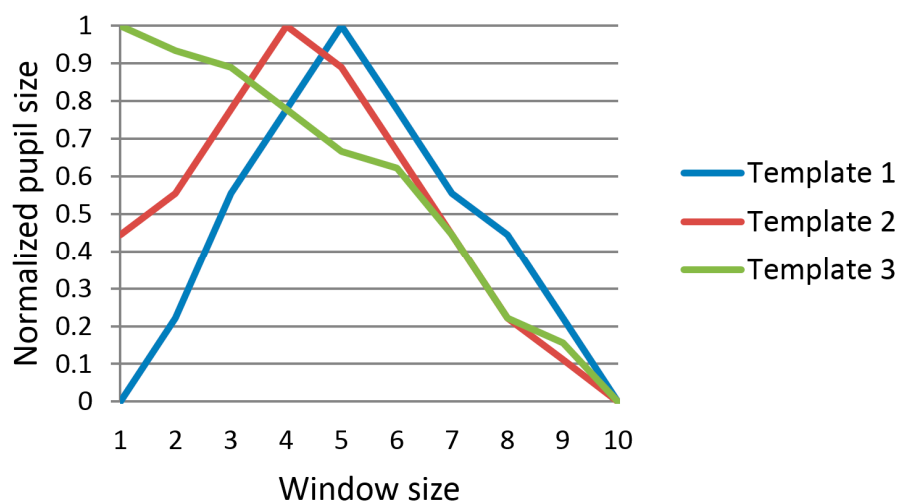

Figure 4. Three template graphs for calculating template matching score.

3.4. Calculating Feature 2 (Change in Gaze Position within Short Dwell Time)

The detected pupil center and glint center (explained in Section 3.2) are used to calculate the gaze position, i.e., feature 2. Initial user calibration is performed to calculate gaze position. For calibration, each user is instructed to examine four positions close to the corners of the monitor [24]. From this, four pairs of pupil centers and glint centers are obtained, as shown in Figure 5.
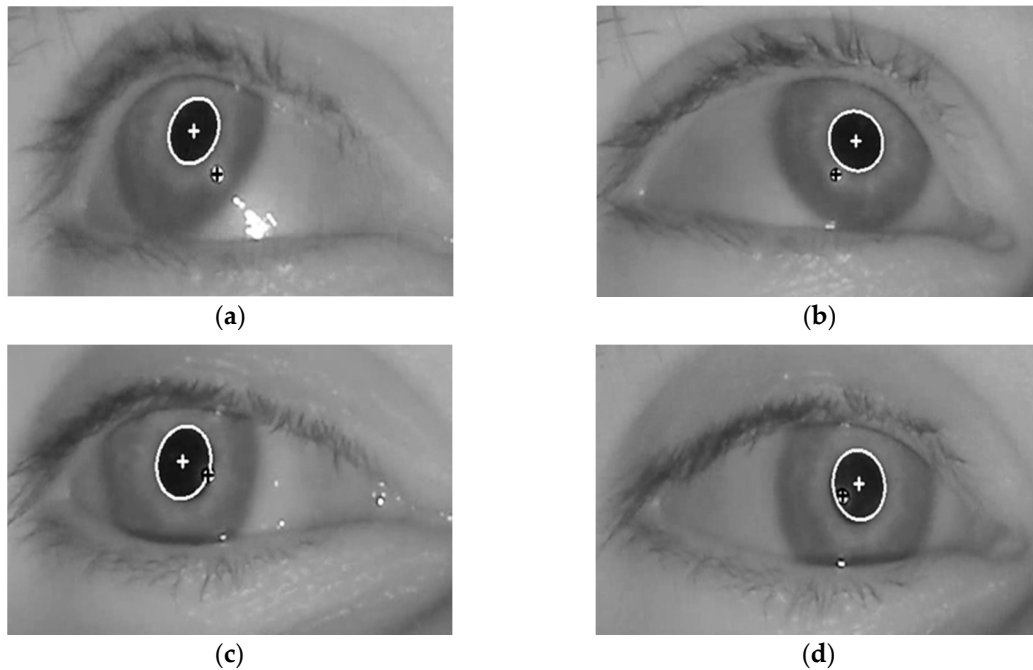
Figure 5. Four images showing the detected centers of glint and the pupil, when a person is looking at the (**a**) upper-left; (**b**) upper-right; (**c**) lower-left; and (**d**) lower-right calibration positions on a monitor.

The position of the center of the pupil is compensated for by that of the glint center, which can reduce the effect of head movement on the variation in gaze position. A geometric transform matrix can be calculated with these four pairs of detected pupil centers and glint centers [24]. This matrix defines the relationship between the region occupied by the monitor and that by the movable pupil, as shown in Figure 6.
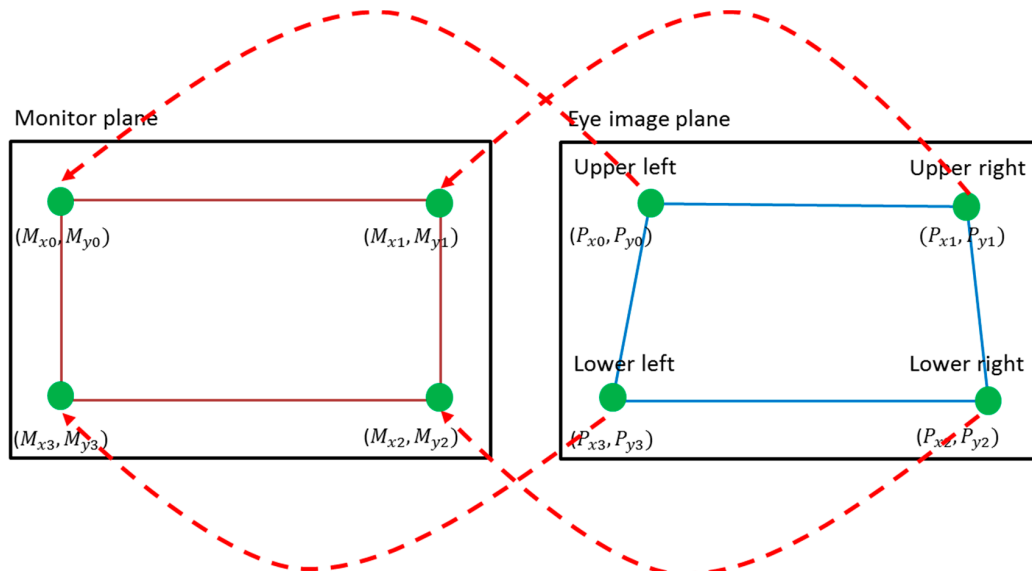

Figure 6. Relationship between pupil movable region (the quadrangle defined by (P_{x0}, P_{y0}), (P_{x1}, P_{y1}), (P_{x2}, P_{y2}), and (P_{x3}, P_{y3})) in the eye image and the monitor region (the quadrangle defined by (M_{x0}, M_{y0}), (M_{x1}, M_{y1}), (M_{x2}, M_{y2}), and (M_{x3}, M_{y3})).

Then, geometric transform matrix is calculated using Equation (5), and the position of the user's gaze (G_x, G_y) is calculated by Equation (6):

$$\begin{bmatrix} M_{x0}\ M_{x1}\ M_{x2}\ M_{x3} \\ M_{y0}\ M_{y1}\ M_{y2}\ M_{y3} \end{bmatrix} = \begin{bmatrix} a & b & c & d \\ e & f & g & h \end{bmatrix} \begin{bmatrix} P_{x0} & P_{x1} & P_{x2} & P_{x3} \\ P_{y0} & P_{y1} & P_{y2} & P_{y3} \\ P_{x0}P_{y0} & P_{x1}P_{y1} & P_{x2}P_{y2} & P_{x3}P_{y3} \\ 1 & 1 & 1 & 1 \end{bmatrix} \tag{5}$$

$$\begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} a & b & c & d \\ e & f & g & h \end{bmatrix} \begin{bmatrix} P\prime_x \\ P\prime_y \\ P\prime_x P\prime_y \\ 1 \end{bmatrix} \tag{6}$$

As the two positions are obtained from the left and the right eyes, the final gaze position is calculated by taking the average of these two gaze positions. Gaze position is not usually moved in case of the user's gaze for target selection. Therefore, the Euclidean distance (Δz_i) of gaze positions between the given (x_i, y_i) and the previous image frames (x_{i-1}, y_{i-1}) is calculated as shown in the Equation (7). Then, feature 2 (change in gaze position within a short dwell time) is calculated from the estimated starting (time) position of the gaze, i.e., S of Equation (8), over a specified short dwell time (the window size W of Equation (8)), and feature 2 becomes smaller in case of the employment of the user's gaze for target selection:

$$\Delta z_i = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2} \tag{7}$$

$$\text{Feature 2 (change in gaze position within short dwell time)} = \sum_{i=S}^{W+S-1} \Delta z_i \tag{8}$$

Although a previous study [24] used a similar concept to measure changes in gaze position, this method involves measuring the change in gaze position by selecting the larger of two changes in gaze position along the horizontal and the vertical directions. For example, if the changes in gaze position are, respectively, 3 and 4 along the horizontal and vertical directions, feature 2 measured by this method is 4 instead of 5 ($\sqrt{3^2 + 4^2}$). By contrast, our method measures the change in gaze position both along the horizontal and the vertical directions, and feature 2 is measured as 5 in this case.

3.5. Calculating Feature 3 (the Texture Information of Monitor Image at Gaze Target)

Baddeley et al. and Yanulevskaya et al. found that edge frequencies are strong contributors to the location of the user's gaze, and have higher correlations to it than other factors [71,72]. Based on this concept, we extract edge-based texture information from the expected gazing location using a Gabor filter [73], and use this information as feature 3 to detect the user's gaze for target selection. The region where the object of interest is located has a larger amount of texture than where it is not. A 2D Gabor filter in the spatial domain is defined as follows:

$$g(x,y) = \left(\frac{1}{2\pi\sigma_x\sigma_y}\right) exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + 2\pi jWx\right] \tag{9}$$

where $g(x,y)$ is defined as the Gabor function along the x- and y-axes. σ_x and σ_y are standard deviations of the function along the x- and y-axes, respectively. W is the radial frequency of a sinusoid. We consider only the real part of the Gabor filter for fast processing. To obtain the Gabor wavelet, $g(x,y)$ in Equation (9) is used as the mother Gabor wavelet. The Gabor wavelet is then obtained by the scaling and rotation of $g(x,y)$ as shown in Equation (10):

$$g_s(x,y) = a^{-m}g\left(x',y'\right), \left(x' = x cos\,\theta + y sin\,\theta,\ y' = -x sin\,\theta + y cos\,\theta\right) \tag{10}$$

where θ is the filter orientation expressed by $\theta = n\pi/K$, K is the number of the filter orientations (n is the integer number), a^{-m} is the filter scale, and $m = 0, \ldots, P$, $P + 1$ is the number of scales. In order to extract the accurate texture, we use 16 Gabor wavelet filters with $K = 4$ and $P = 3$, as shown in Figure 7. The average number of magnitudes obtained using Gabor wavelet filters within the ROI is used as feature 3. For this, the ROI for the application of the Gabor filter is defined based on the user's gaze as shown in Figure 8. According to the gaze position, the magnitude of texture in the ROI varies, and the user's gaze shows a high value for feature 3, due to the complex texture of the monitor image.
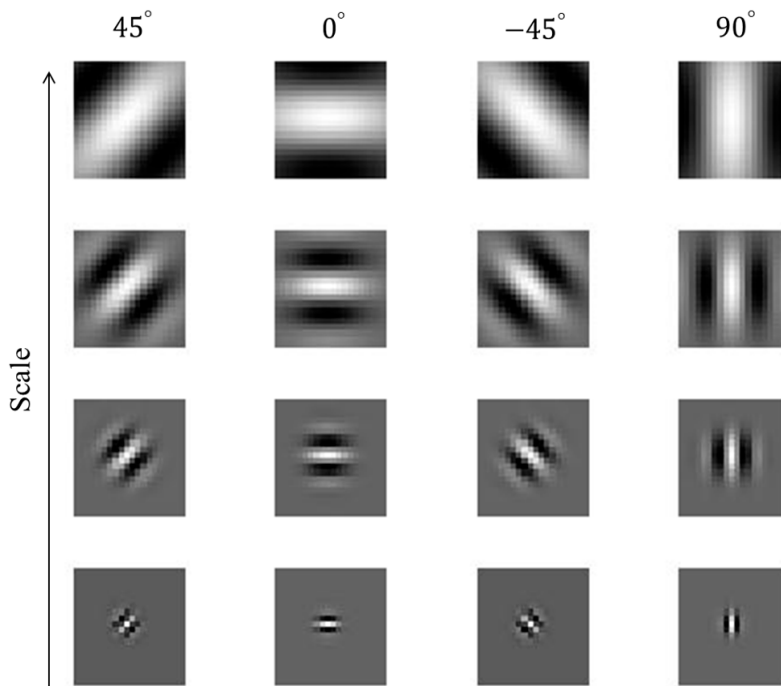

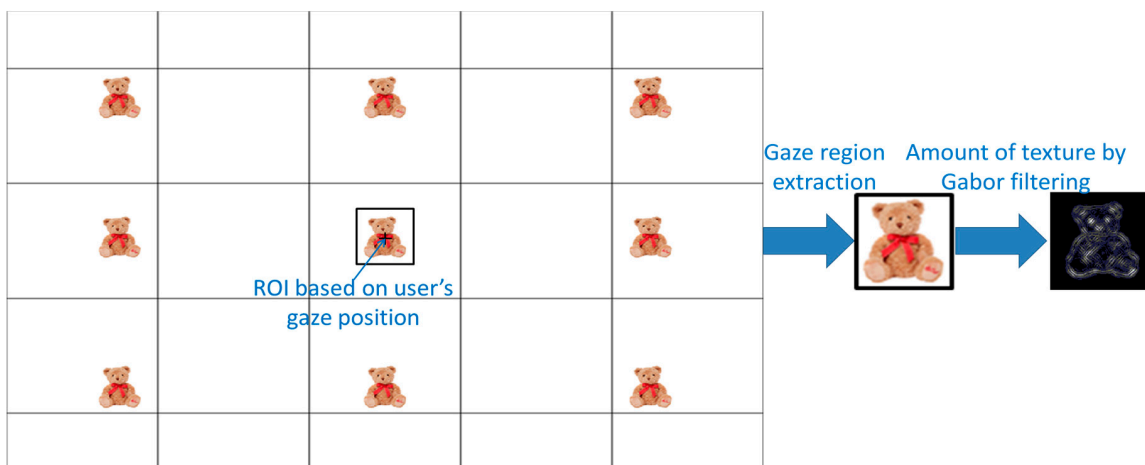**Figure 7.** 16 Gabor filters with four scales and four orientations.

Figure 8. ROI for applying Gabor filter based on the user's gaze position.

3.6. Fuzzy Logic System for Detecting User's Gaze for Target Selection

3.6.1. Explanation of Fuzzy Membership Functions

To detect the user's gaze for target selection, our method uses a fuzzy system with three input features, i.e., features 1–3 (explained in Sections 3.3–3.5), as shown in Figure 9. As explained in

Sections 3.3–3.5, features 1 and 2 are smaller, whereas feature 3 is larger in case the user's gaze is used. Through normalization based on min-max scaling, these three features are made to range from 0 to 1. To make these three features consistent with one another, features 1 and 2 are recalculated by subtracting them from the maximum value (1). Therefore, features 1–3 are larger in when the user's gaze is used, and are used as inputs to the fuzzy logic system. Based on the output of the fuzzy system, we can determine whether the user is gazing at the selected target.
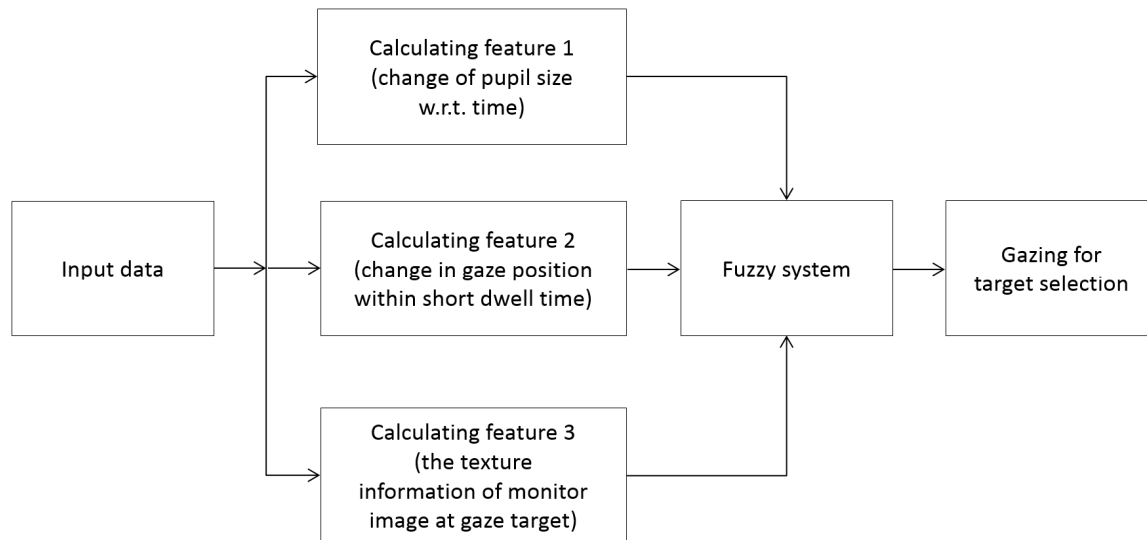

Figure 9. Fuzzy method for detecting user's gaze for target selection.

Figure 10 shows the membership functions for the three input features 1–3. The input values were classified into two classes in the membership function: Low (L) and High (H). In general, these value classes are not separated, and the membership functions are defined to have overlapping areas, as shown in Figure 10. With a small number of input data items, we obtained the distributions of features 1–3 as shown in Figure 10 and, based on the maximum entropy criterion [74], designed the input membership functions. For fair experiments, these data were not used for all experiments reported in Section 4.

We first define the rough shape of the input membership functions as linear by considering processing speed and the complexity of problem because such functions have been widely used in fuzzy applications [75–77]. The defined input membership functions are as follows:

$$\mu_{L\_feature\ i}(x) = \left\{ \begin{array}{ll} 1 & \text{for } 0 \leq x \leq a_{L\_i} \\ p_{L\_i}x + q_{L\_i} & \text{for } a_{L\_i} \leq x \leq b_{L\_i} \\ 0 & \text{for } b_{L\_i} \leq x \leq 1 \end{array} \right\} \tag{11}$$

$$\mu_{H\_feature\ i}(x) = \left\{ \begin{array}{ll} 0 & \text{for } 0 \leq x \leq a_{H\_i} \\ p_{H\_i}x + q_{H\_i} & \text{for } a_{H\_i} \leq x \leq b_{H\_i} \\ 1 & \text{for } b_{H\_i} \leq x \leq 1 \end{array} \right\} \tag{12}$$

In Equations (11) and (12), i = 1, 2, and 3. $\mu_{L\_feature\ i}(x)$ is the L membership function of feature i whereas $\mu_{H\_feature\ i}(x)$ is its H membership function. Then, we can obtain the following equations:

$$p_{L\_feature\ i} = \sum_{x=0}^{1} m_{L\_feature\ i}(x)\mu_{L\_feature\ i}(x) \tag{13}$$

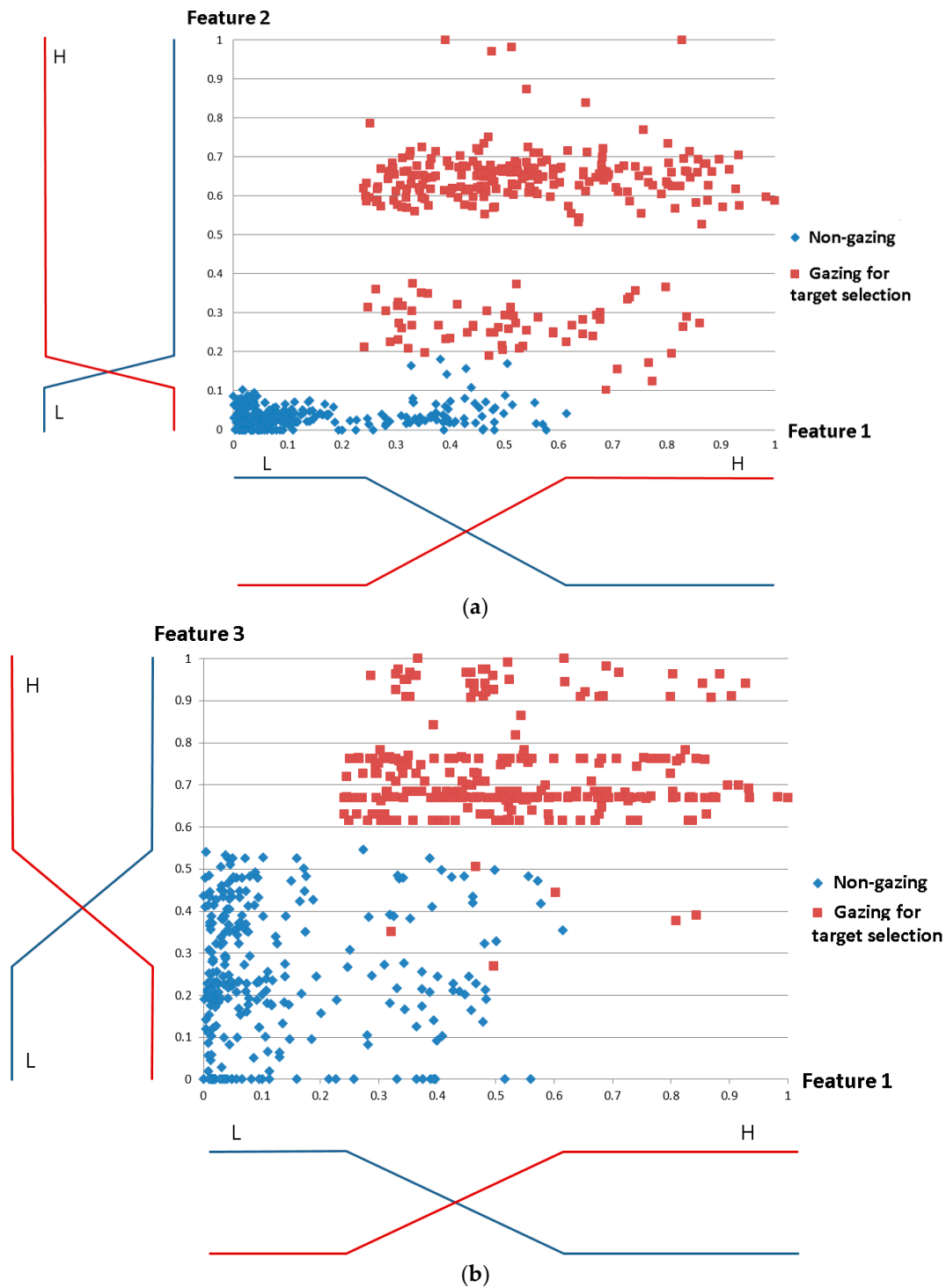$$p_{H\_feature\ i} = \sum_{x=0}^{1} m_{H\_feature\ i}(x)\mu_{H\_feature\ i}(x) \tag{14}$$

Figure 10. Designing optimal input fuzzy membership functions for features 1–3 based on maximum entropy criterion. (**a**) Membership functions for features 1 and 2; (**b**) Membership functions for features 1 and 3.

In Equations (13) and (14), i = 1, 2, and 3. $\mu_{L\_feature\ i}(x)$ is the L membership function of feature i of Equation (11), whereas $\mu_{H\_feature\ i}(x)$ is the H membership function of feature i in Equation (12). In addition, $m_{L\_feature\ i}(x)$ is the L (data) distribution of feature i (non-gazing data of Figure 10), whereas $m_{H\_feature\ i}(x)$ is the H (data) distribution of feature i (gazing data for target selection of Figure 10). Based on Equations (13) and (14), the entropy can be calculated as follows:

$$H(a_{L\_i}, b_{L\_i}, p_{L\_i}, q_{L\_i}, a_{H\_i}, b_{H\_i}, p_{H\_i}, q_{H\_i}) = -p_{L_{featurei}} \log\left(p_{L_{featurei}}\right) - p_{H_{featurei}} \log\left(p_{H_{featurei}}\right) \quad (15)$$

where i = 1, 2, and 3. Based on the maximum entropy criterion [74], the optimal parameters of $(a_{L\_i}, b_{L\_i}, p_{L\_i}, q_{L\_i}, a_{H\_i}, b_{H\_i}, p_{H\_i}, q_{H\_i})$ of feature i are obtained by being chosen when the entropy $H(a_{L\_i}, b_{L\_i}, p_{L\_i}, q_{L\_i}, a_{H\_i}, b_{H\_i}, p_{H\_i}, q_{H\_i})$ is maximized. From this, we can obtain the input membership functions of features 1–3.

These membership functions are used to convert input values into degrees of membership. In order to determine whether the user's gaze for target selection occurs, the output value is also described in the form of a linear function from the membership functions, as in Figure 11 that shows the three functions of L, M, and H. Using these output membership functions, the fuzzy rule table, and a combination of the defuzzification method with the MIN and MAX rules, the optimal output value can be obtained (see details in Section 3.6.3)
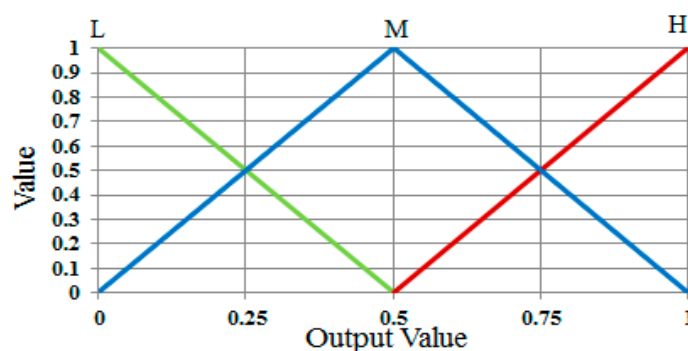

Figure 11. Definitions of output membership functions.

3.6.2. Fuzzy Rules Based on Three Input Values

As explained in Section 3.6.1 and Figure 10, the values of features 1–3 are larger in case the user's gaze is employed for target selection. We define the output of our fuzzy system as "H" in case it is, and as "L" when it is not. Based on this, we define the fuzzy rules shown in Table 2.

Table 2. Fuzzy rules based on features 1–3.

| Feature 1	Feature 2	Feature 3	Output of Fuzzy System
L	L	L	L
L	L	H	L
L	H	L	M
L	H	H	H
H	L	L	L
H	L	H	M
H	H	L	H
H	H	H	H

3.6.3. Determining the User's Gaze for Target Selection Based on Defuzzification Methods

Using the three normalized input features, six corresponding values can be acquired using the input membership functions as shown in Figure 12. Three functions are defined as $g_{f1}(\cdot)$, $g_{f2}(\cdot)$, and $g_{f3}(\cdot)$. The corresponding output values of the three functions with input values of $f1$(feature 1), $f2$(feature 2), and $f3$(feature 3) are denoted by (g_{f1}^L, g_{f1}^H), (g_{f2}^L, g_{f2}^H), and (g_{f3}^L, g_{f3}^H). For example, suppose that the three input values for $f1, f2$, and $f3$ are 0.30, 0.50, and 0.45, respectively, as shown in Figure 12. The values of (g_{f1}^L, g_{f1}^H), (g_{f2}^L, g_{f2}^H), and (g_{f3}^L, g_{f3}^H) are (0.75(L), 0.25(H)), (0.00(L), 1.00(H)),
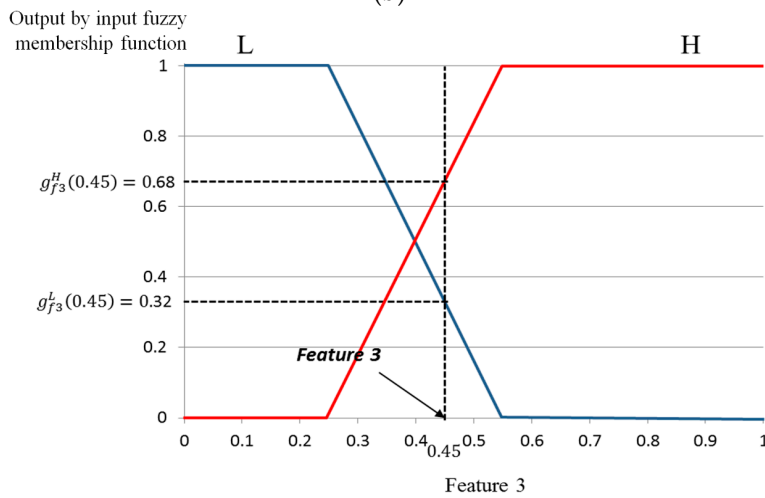
and (0.32(L), 0.68(H)), respectively, as shown in Figure 12. With these values, we can obtain the following eight combinations: (0.75(L), 0.00(L), 0.32(L)), (0.75(L), 0.00(L), 0.68(H)), (0.75(L), 1.00(H), 0.32(L)) . . . , and (0.25(H), 1.00(H), 0.68(H)).
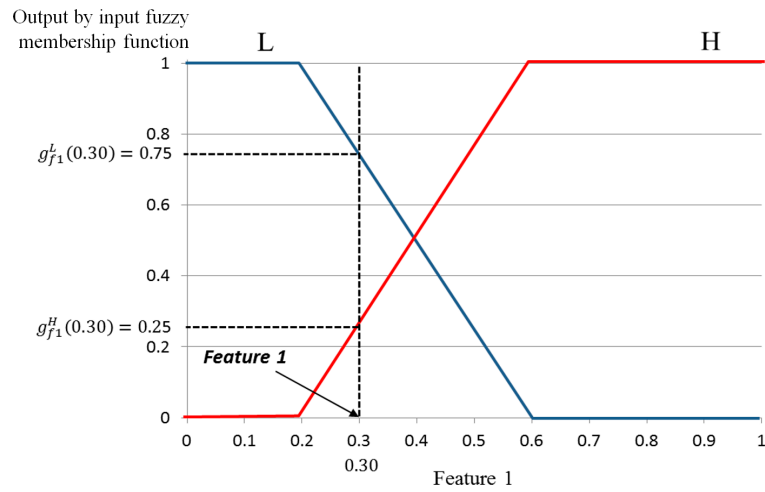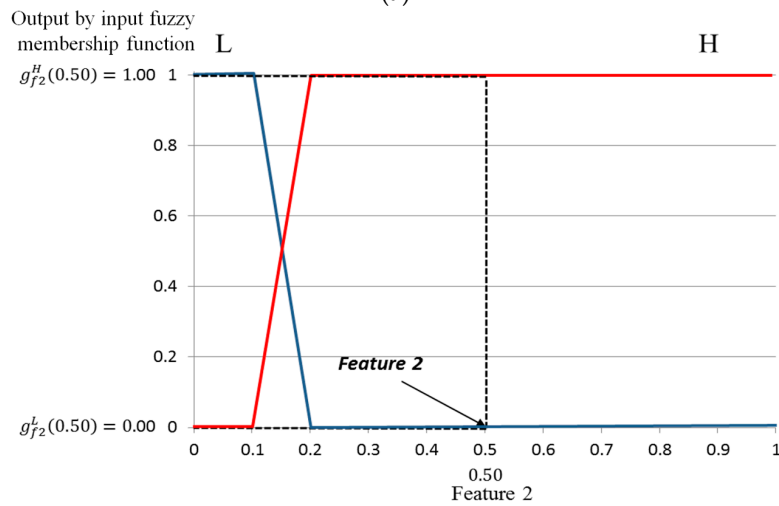
Figure 12. Finding the output value of the input membership function for three features: (**a**) feature 1; (**b**) feature 2; and (**c**) feature 3.

The proposed method defines which of L and H can be used as inputs for the defuzzification step using the eight rules in Table 2. The MIN or MAX method is commonly used for this purpose. In the MIN method, the minimum value is selected from each combination set of three members and used as input for defuzzification. For the MAX method, the maximum value is selected and used as defuzzification input.

For example, for a combination set of (0.25(H), 0.00(L), 0.68(H)), the MIN method selects the minimum value (0.00) as input. For the MAX method, the maximum value (0.68) is selected. Then, on the basis of fuzzy logic rules from Table 2 (if H, L, and H, then M), values of 0.00(M) and 0.68(M) are finally determined by the MIN and MAX methods, respectively.

All values calculated by the MIN and MAX rules with eight combinations of this example are listed in Table 3. We refer to these values as "inference values" (IVs). As shown in Table 3, these IVs are used as inputs for defuzzification to obtain the output. The MIN and MAX rules were compared in our experiments.

Table 3. IVs obtained with eight combinations.

Feature 1	Feature 2	Feature 3	IV	
			MIN Rule	**MAX Rule**
0.75(L)	0.00(L)	0.32(L)	0.00(L)	0.75(L)
0.75(L)	0.00(L)	0.68(H)	0.00(L)	0.75(L)
0.75(L)	1.00(H)	0.32(L)	0.32(M)	1.00(M)
0.75(L)	1.00(H)	0.68(H)	0.68(H)	1.00(H)
0.25(H)	0.00(L)	0.32(L)	0.00(L)	0.32(L)
0.25(H)	0.00(L)	0.68(H)	0.00(M)	0.68(M)
0.25(H)	1.00(H)	0.32(L)	0.25(H)	1.00(H)
0.25(H)	1.00(H)	0.68(H)	0.25(H)	1.00(H)

Several defuzzification methods are compared on fuzzy systems. We considered five of them, i.e., first of maxima (FOM), last of maxima (LOM), middle of maxima (MOM), center of gravity (COG), and bisector of area (BOA) [76–79]. In each defuzzification method excluding COG and BOA, the maximum values of IVs were used to calculate the output value. The maximum IVs were $IV_1(L)$ and $IV_2(M)$, as shown in Figure 13a. In the FOM defuzzification method, as the name implies, the first value after defuzzification is selected as the optimal weight value, and is represented as w_1 in Figure 13a. The last defuzzification value, i.e., w_3 is the optimal weight value selected by the LOM method. The method that shows that the optimal weight value is obtained from the middle value of the optimal weight values of FOM and LOM is referred to as the MOM method, i.e., $(w_{\mathrm{MOM}} = \frac{1}{2}(w_1 + w_3))$.

The output scores for the methods using the COG and the BOA are calculated differently from those of other defuzzification methods. The COG method is also known as center-of-area defuzzification. It calculates the output score based on the geometrical center of the non-overlapping regions, and is formed by the regions defined by all IVs. As shown in Figure 13b, regions R_1, R_2, and R_3 are defined based on all IVs, where region R_1 is the quadrangle defined by connecting four points $(0, IV_1(L))$, $(w_1, IV_1(L))$, $(0.5, 0)$, and $(0, 0)$, R_2 is the quadrangle defined by connecting four points $(w_2, IV_2(M))$, $(w_3, IV_2(M))$, $(1, 0)$, and $(0, 0)$, and R_3 is that defined by connecting four points $(w_4, IV_3(H))$, $(1, IV_3(H))$, $(1, 0)$, and $(0.5, 0)$.

Finally, the optimal weight value of the fuzzy system (w_5) is calculated from the center of gravity of regions R_1, R_2, and R_3 (considering the overlapping regions of R_1, R_2, and R_3 only once), as shown in Figure 13b. The BOA is calculated by a vertical line (w_6) that divides the region defined by the IVs into two sub-regions of equal area, which is why it is called the bisector method. Following this, our fuzzy system determines that user gazes at the position for target selection, if the output score of the fuzzy system is greater than a threshold defined for the fuzzy system.
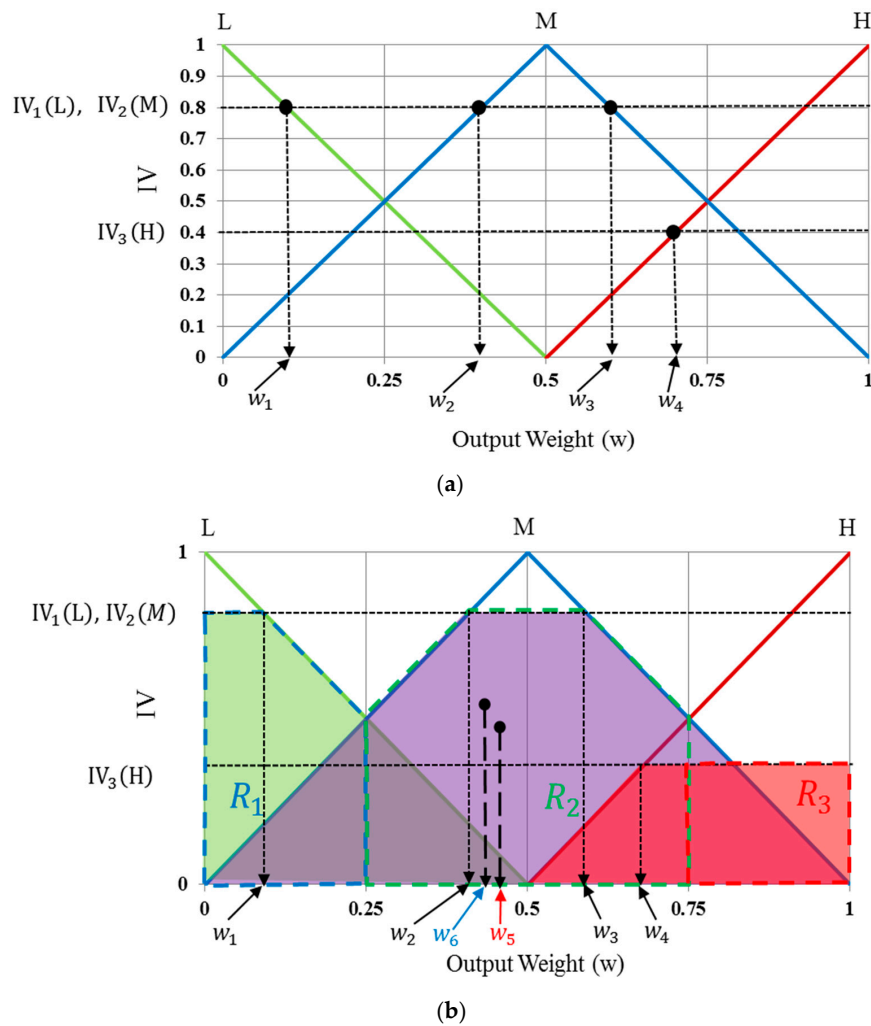
(a)

(b)

Figure 13. Obtaining output score values of fuzzy system using various defuzzification methods. (**a**) FOM, LOM, and MOM; (**b**) COG and BOA.

4. Experimental Results

The performance of the proposed method for detecting user gaze for target selection was measured through experiments performed with 15 participants, where each participant attempted two trials for each of three target objects, i.e., a teddy bear, a bird, and a butterfly, displayed at nine positions on a 19-inch monitor, as shown in Figure 14. That is, three experiments were performed with these objects (teddy bear, bird, and butterfly) with each at nine positions on the screen. We collected 270 data items (15 participants × 2 trials × 9 gaze positions) for gazing for target selection, i.e., true positive (TP) data, and the same number of non-gazing data, i.e., true negative (TN) data, for each of the three experiments. Most participants were graduate students, and some of them were faculty or staff members of our university's department. They were randomly selected by considering the variation in eye characteristics with age, gender, and nationality. All participants voluntarily participated in our experiments. Of the 15, five wore glasses and four people wore contact lenses. The remained 6 people did not wear glasses and contact lens. The ages of the participants ranged from 20 s to 40 s (mean age 29.3 years). Nine participants were male, and the other 6 ones were female. We made sure that participants of different nationalities were involved in our experiments: one Mongolian, one Tanzanian, two Pakistanis, four Vietnamese, and seven Koreans. Before the experiments, we provided the sufficient explanations of our experiments to all participants, and obtained informed (written) consents from all the participants.
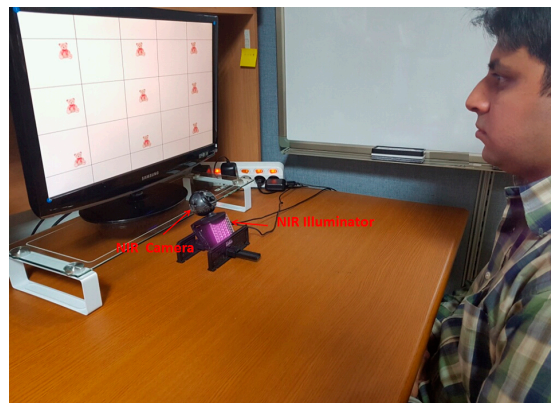
(a)
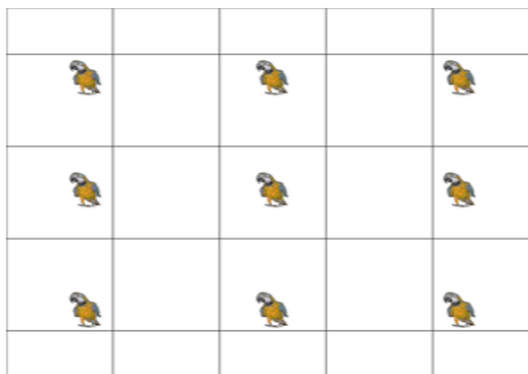
(b)

Figure 14. Experimental setup for proposed method. (**a**) Example of experimental environment; (**b**) Three screens of teddy bear (upper), bird (middle), and butterfly (lower) used in our experiments.

In the first experiment, we compared the accuracy of our method, in detecting the boundary between the pupil and the glint as well as the center of each, with a previous method [24]. As shown in Figure 15b,c, the boundary and center of the pupil detected by our method were closer to the ground truth than those calculated by the previous method. Moreover, as shown in Figure 15d,e, the detected boundary and center of the glint according to our method were closer to the ground truth than the previous method. In this experiment, the boundary and center of the ground truth were manually chosen. Further, for all images, we measured detection errors based on Euclidean distance between the center of the ground truth and the center as detected by our method and the previous method [24]. As shown in Table 4, our method yielded a higher accuracy (lower error) in detecting the centers of the pupil and the glint than the previous method.
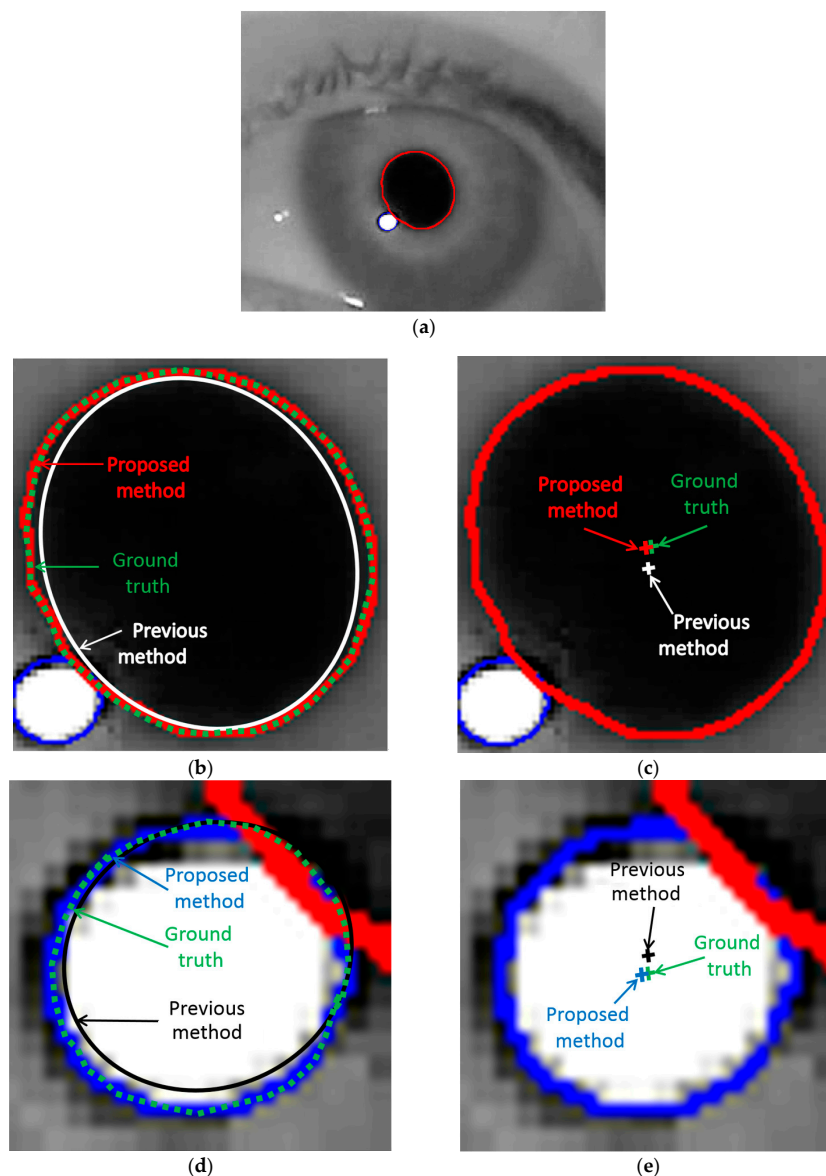

Figure 15. Comparative examples of detection of boundaries and centers of pupil and glint by our method, and the previous method and the ground truth. (**a**) Detected boundaries of pupil and glint in eye image; (**b**) Comparison of detection of boundaries of pupil; (**c**) Comparison of detection of center of pupil; (**d**) Comparison of detection of boundaries of glint; (**e**) Comparison of detection of center of glint.

Table 4. Average Euclidean distance between the center of the ground truth the detected center by our method and previous method (unit: pixels).

Pupil/Glint Center	Method	Average Euclidean Distance
Pupil Center	Proposed method	2.1
	Previous method [24]	4.6
Glint Center	Proposed method	2.5
	Previous method [24]	4.7

In the second experiment, the accuracy of the detection of the user's gaze for target selection on TP and TN data, according to various defuzzification methods, was compared in terms of equal error rate (EER). Two types of errors, types I and II, were considered. The error of incorrectly classifying TP data as TN data was defined as a type I error, and that of incorrectly classifying TN data as TP data was defined as a type II error. As explained in the previous section, our system determines that the user's gaze is employed (TP) if the output score of the fuzzy system is higher than the threshold. If not, our system determines that user's gaze is not employed (TN). Therefore, type I and II errors occur depending on the threshold. If the threshold is increased, type 1 error increases and type II error decreases. With a smaller threshold, type I error decreases whereas type II error increases. When type I and II errors are most similar to the appropriate threshold, the EER is calculated by averaging the two errors.

Experiment 1 (Bear)

In the first experiment, a teddy bear was used as target object, as shown in Figure 14b. The classification results of TP and TN data according to the five defuzzification methods that used the MIN and MAX rules are listed in Tables 5 and 6, respectively. As indicated in these tables, the smallest EER (approximately 0.19%) of classification was obtained by the COG with the MIN rule.

Table 5. Type I and II errors with EER using the MIN rule (unit: %).

Defuzzification Method	Threshold	No. of Type I Errors	No. of Type II Errors	EER
FOM	0.36	9	6	2.78
MOM	0.50	11	0	2.04
LOM	0.73	12	16	5.19
BOA	0.27	1	1	0.37
COG	0.30	0	1	0.19

Table 6. Type I and II errors with EER using the MAX rule (unit: %).

Defuzzification Method	Threshold	No. of Type I Errors	No. of Type II Errors	EER
FOM	0.09	111	98	38.7
MOM	0.50	151	0	27.96
LOM	0.89	126	124	46.3
BOA	0.50	110	98	38.52
COG	0.50	101	79	33.33

Figures 16 and 17 show the receiver operating characteristic (ROC) curves for the classification results of TP and TN data according to the various defuzzification methods using MIN or MAX rules, respectively. The ROC curve represents the change in type I error (%) according to the increase in 100—type II error (%).
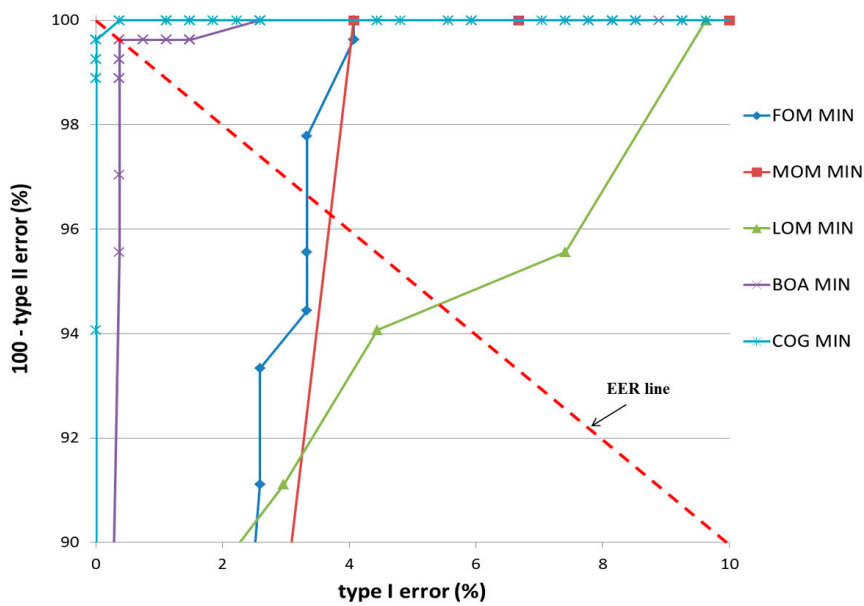
Figure 16. ROC curves of the classification results of TP and TN data according to different defuzzification methods with the MIN rule.
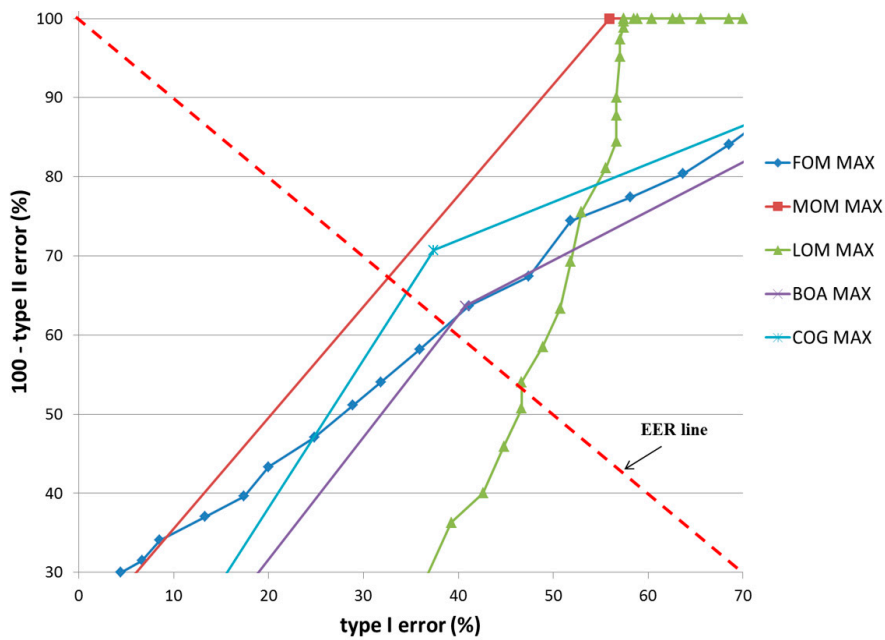

Figure 17. ROC curves of the classification results of TP and TN data according to different defuzzification methods with the MAX rule.

In case that type I and II errors were small, the accuracy of that particular method was regarded as high. Therefore, if the ROC curve was closer to (0, 100) ("type I error" of 0% and "100—type II error" of 100%) on the graph, its accuracy was regarded as higher. As shown in these figures, the accuracy of classification by the COG with the MIN rule was higher than obtained by other defuzzification methods.

Experiment 2 (Bird)

In the second experiment, a bird was used as target object as shown in Figure 14b. The classification results of TP and TN data according to the five defuzzification methods that used the MIN and MAX rules are listed in Tables 7 and 8, respectively. As indicated in these tables, the smallest EER (approximately 0%) of classification was obtained by the COG with the MIN rule.

Table 7. Type I and II errors with EER using the MIN rule (unit: %).

Defuzzification Method	Threshold	No. of Type I Errors	No. of Type II Errors	EER
FOM	0.26	7	8	2.78
MOM	0.50	14	0	2.60
LOM	0.60	19	21	7.41
BOA	0.25	0	1	0.19
COG	0.27	0	0	0

Table 8. Type I and II errors with EER using the MAX rule (unit: %).

Defuzzification Method	Threshold	No. of Type I Errors	No. of Type II Errors	EER
FOM	0.11	142	140	52.22
MOM	0.50	162	0	30
LOM	0.91	90	87	32.78
BOA	0.50	101	71	31.85
COG	0.50	85	75	29.63

Figures 18 and 19 show the ROC curves for the classification results of TP and TN data according to the various defuzzification methods using MIN or MAX rules. As is shown, the accuracy of classification by the COG with the MIN rule was higher than for other defuzzification methods.
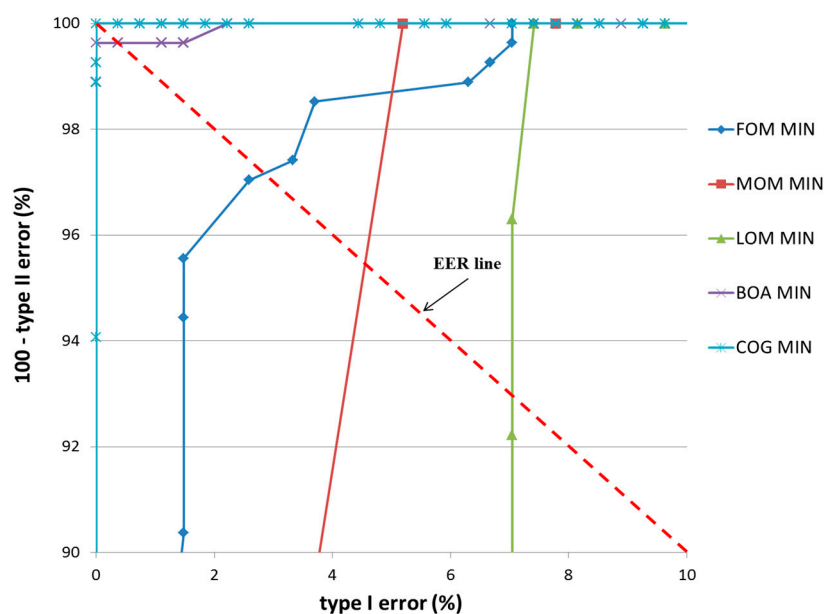

Figure 18. ROC curves of the classification results of TP and TN data according to different defuzzification methods with the MIN rule.
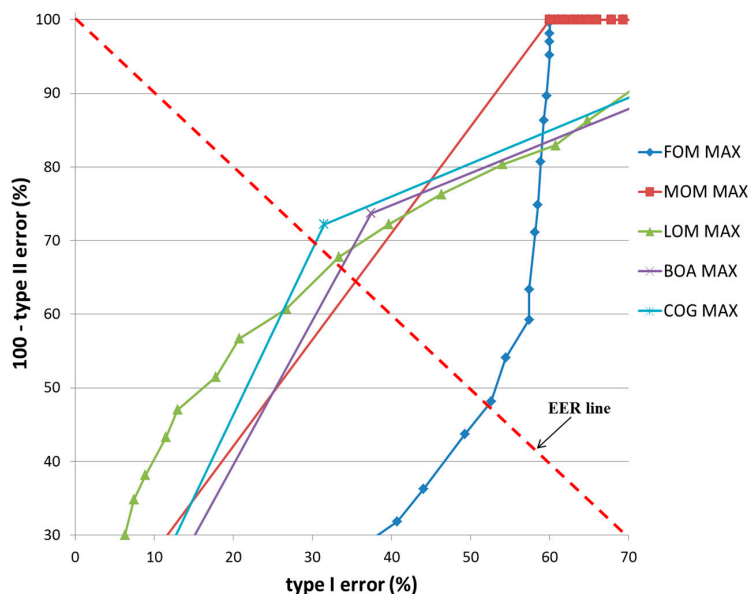
Figure 19. ROC curves of the classification results of TP and TN data according to different defuzzification methods with the MAX rule.

Experiment 3 (Butterfly)

In the third experiment, a butterfly was used as target object as shown in Figure 14b. The classification results of TP and TN data according to the five defuzzification methods that used the MIN and MAX rules are listed in Tables 9 and 10, respectively. As indicated in these tables, the smallest EER (approximately 0.19%) of classification was obtained by the COG with the MIN rule.

Table 9. Type I and II errors with EER using the MIN rule (unit: %).

Defuzzification Method	Threshold	No. of Type I Errors	No. of Type II Errors	EER
FOM	0.35	12	13	4.63
MOM	0.47	12	0	2.22
LOM	0.71	13	15	5.19
BOA	0.37	2	0	0.37
COG	0.30	0	1	0.19

Table 10. Type I and II errors with EER using the MAX rule (unit: %).

Defuzzification Method	Threshold	No. of Type I Errors	No. of Type II Errors	EER
FOM	0.11	151	142	54.26
MOM	0.50	184	0	34.07
LOM	0.90	81	94	32.41
BOA	0.50	135	75	38.89
COG	0.50	120	89	38.7

Figures 20 and 21 show the ROC curves for the classification results of TP and TN data according to the various defuzzification methods using the MIN or MAX rules. As is shown, the accuracy of classification by the COG with the MIN rule was higher than by other defuzzification methods.

The above results were verified by comparing the proposed method using three features (change of pupil size w.r.t. time by template matching, change in gaze position within short dwell time, and the texture information of monitor image at gaze target) with the previous method [24] using two features (change of pupil size w.r.t. time by peakedness, and change in gaze position within short dwell time). In addition, we compared the accuracy by proposed method with that by using three features (change

of pupil size w.r.t. time by peakedness, change in gaze position within short dwell time, and the texture information of monitor image at gaze target). For convenience, we call the last method "Method A". The only difference between the proposed method and "Method A" is that change in pupil size w.r.t. time is measured by template matching in proposed method but by peakedness in "Method A".
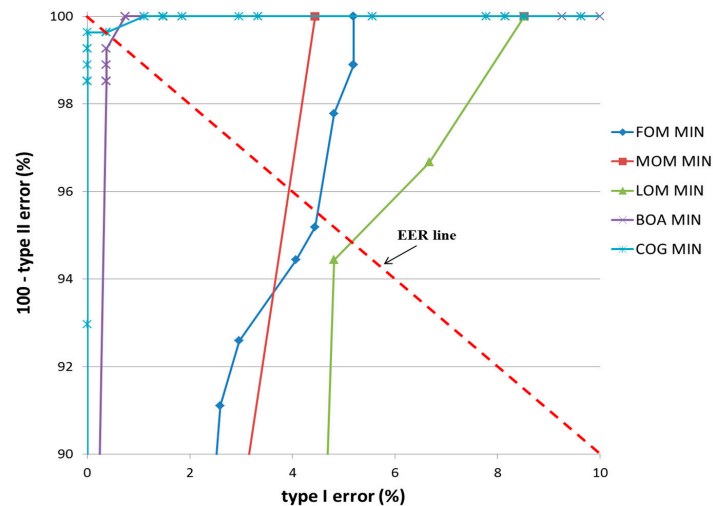

Figure 20. ROC curves of the classification results of TP and TN data according to different defuzzification methods with the MIN rule.
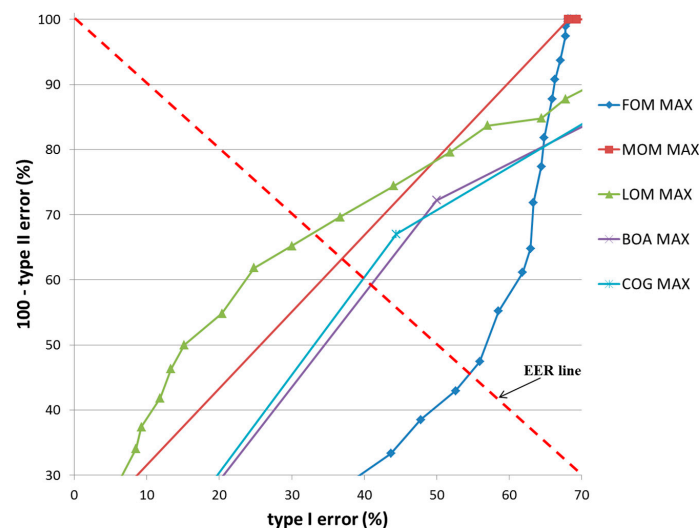

Figure 21. ROC curves of the classification results of TP and TN data according to different defuzzification methods with the MAX rule.

In all cases, the ROC curves of the highest accuracy among the various defuzzification methods with the MIN or MAX rules are shown. The accuracy of by proposed method was always higher than that of "Method A" and the previous method [24] in all cases of the three experiments, i.e., bear, bird, and butterfly, as shown in Figure 22.

In the next experiment, we compared the accuracy of proposed method, "Method A", and the previous method [24] when noise was included in the input data. All features of the proposed and the previous method was affected by the accurate detection of pupil size and gaze position, which were in turn affected by the performance of the gaze tracking system. Thus, we included Gaussian random noises in the detected pupil size and gaze position.
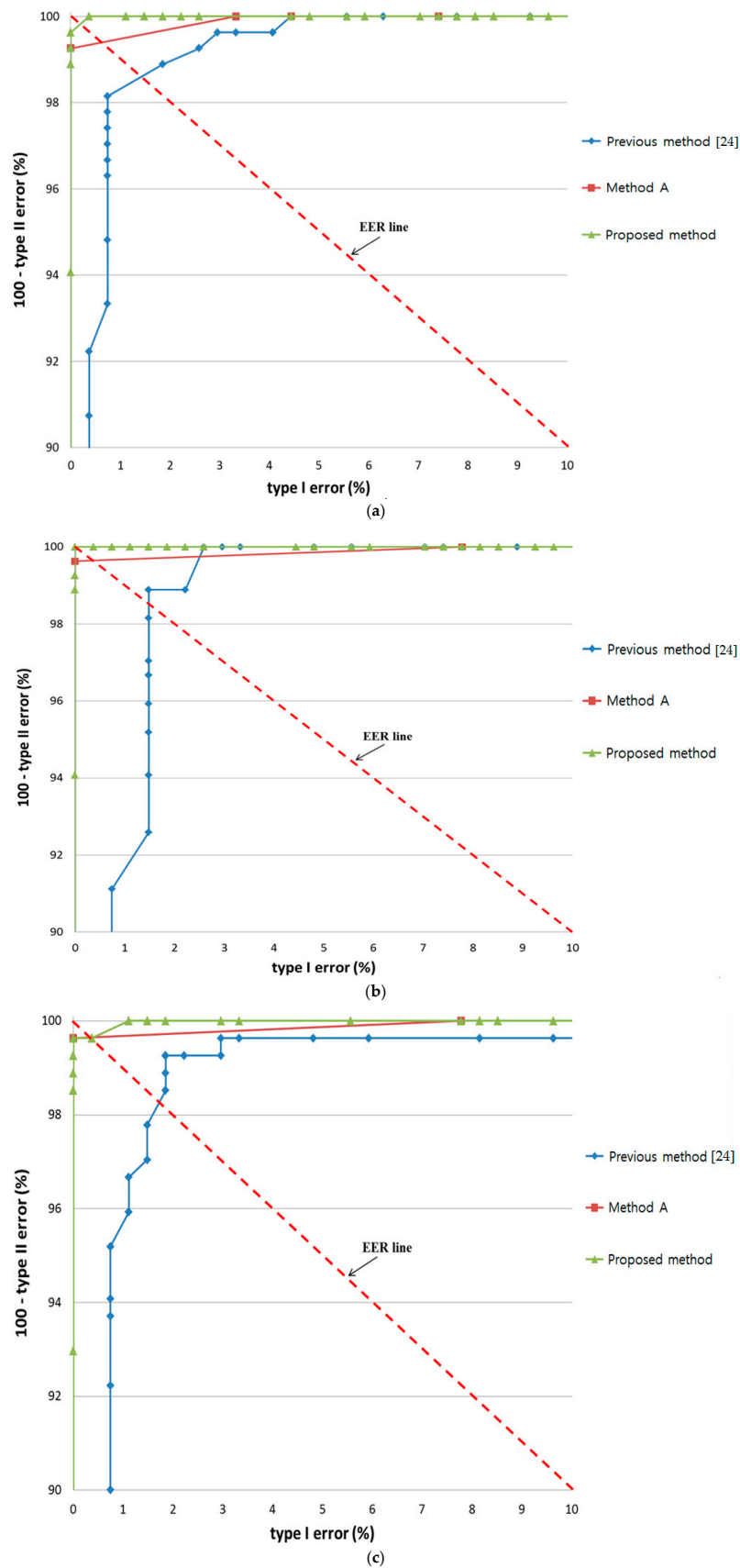
Figure 22. Comparison of results by proposed method with the results by "Method A" and previous method with experiments involving (**a**) bear; (**b**) bird; and (**c**) butterfly.

As shown in Figure 23, noise had a stronger effect on the previous method [24] and "Method A" than the proposed method in all three experiments. A notable decrease in accuracy with increase in EER was observed in the latter two methods when noise caused incorrect detection of pupil size and gaze.
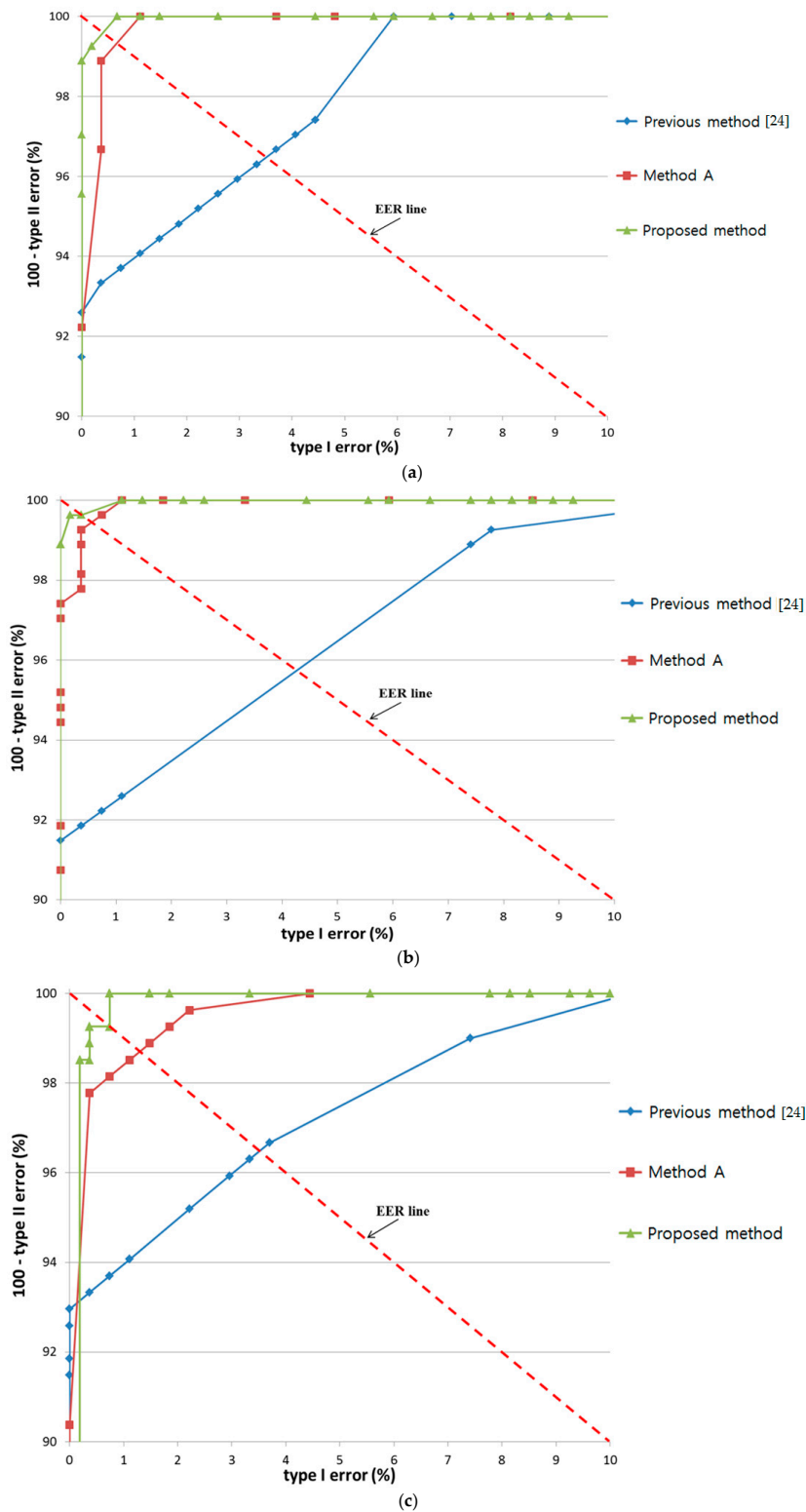

Figure 23. Effect of noise on results by proposed method with the results by "Method A" and previous method with experiments involving (**a**) bear; (**b**) bird; and (**c**) butterfly.

In the next experiments, we compared the usability of our system with the conventional dwell time-based selection method [13,14]. We requested the 15 participants to rate the level of convenience and interest in performing the target selection task of the proposed method and a dwell time-based method by using a questionnaire (5: very convenient, 4: convenient, 3: normal, 2: inconvenient, 1: very inconvenient, in case of a convenience questionnaire) (5: very interesting, 4: interesting, 3: normal, 2: uninteresting, 1: very uninteresting, in case of an interest questionnaire). In order to render our results unaffected by participant learning and physiological state, such as fatigue, we provided a rest time of 10 minutes to each participant between experiments. Based on [13,14], dwell time for target selection was set at 500 ms. That is, when the change in the user's gaze position for our feature 2 was lower than the threshold, and this state was maintained for longer than 500 ms, target selection was activated.

The average scores are shown in Figure 24, which shows that our method scored higher than the conventional dwell time-based method in terms of both convenience and interest.
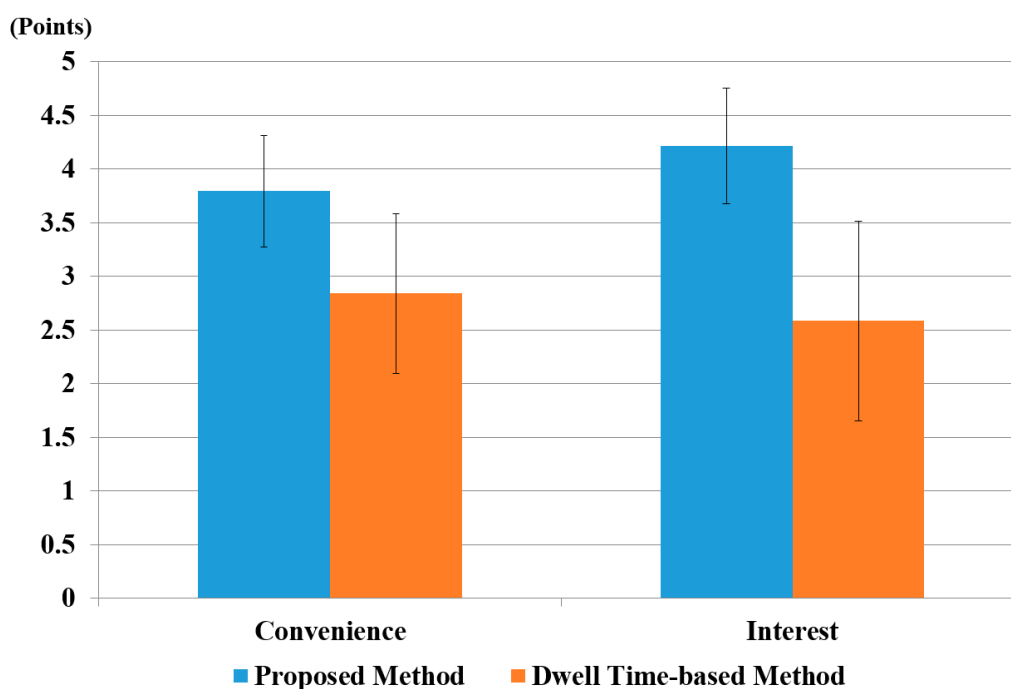

Figure 24. The comparative results of subjective tests of 15 users using the dwell time-based method and the proposed method.

We also performed a *t*-test [80] to prove that user's convenience on the proposed method was statistically higher than that of the conventional dwell time-based method. The *t*-test was performed by taking two independent samples of data: user's convenience on our system ($\mu = 3.8$, $\sigma = 0.5$) and with the conventional dwell time-based method ($\mu = 2.8$, $\sigma = 0.7$). The calculated *p*-value was approximately 3.4×10^{-4}, which was smaller than the 99% (0.01) significance level. Hence, the null hypothesis for the *t*-test, i.e., no difference between the two independent samples, was rejected. Therefore, we can conclude that there was a significant difference, up to 99%, in user convenience between our proposed method and the dwell time-based method.

A *t*-test analysis based on user interest was also performed. This also yielded similar results, i.e., user interest on our proposed system ($\mu = 4.2$, $\sigma = 0.5$) was higher than with the dwell time-based system ($\mu = 2.6$, $\sigma = 0.9$). The calculated *p*-value was 6.74×10^{-6}, i.e., smaller than the significance level of 99% (0.01). Therefore, we concluded a significant difference in user interest between our system and the dwell time-based system. The average score for convenience and interest was higher for the proposed method because it is more natural than a conventional dwell time-based method.

In order to analyze the effect of the difference in size between two groups, we performed Cohen's *d* analysis [81]. It classifies the difference as small if it is within the range 0.2–0.3, medium if it is 0.5, and large if it is greater than or equal to 0.8. We calculated the value of Cohen's *d* for convenience and interest. For user convenience, it was 1.49, which was in the large-effect category; hence, it had a major effect on the difference between the two groups. For user interest, the calculated Cohen's *d* was approximately 2.14, which was also in the large-effect category. Hence, we concluded that user convenience and user interest showed a large effect in the difference between our proposed method and dwell time-based method.

To confirm the practicality of our method, we performed additional experiments using an on-screen keyboard based on our method, where each person typed a word through our system on an on-screen keyboard displayed on a monitor, as shown in Figure 25. All 15 subjects from before participated in the experiments, and a monitor with a 19-inch screen and resolution of 1680 × 1050 pixels was used. Twenty sample words were selected based on frequency of use as shown in Table 11 [82]. As shown in Table 11, the left-upper words "the" and "and" were more frequently used than the right-lower words "all" and "would". If the user's gaze detected by our method was associated with a specific button, and our method had determined the user's gaze, the corresponding character was selected and displayed, as shown in Figure 25.
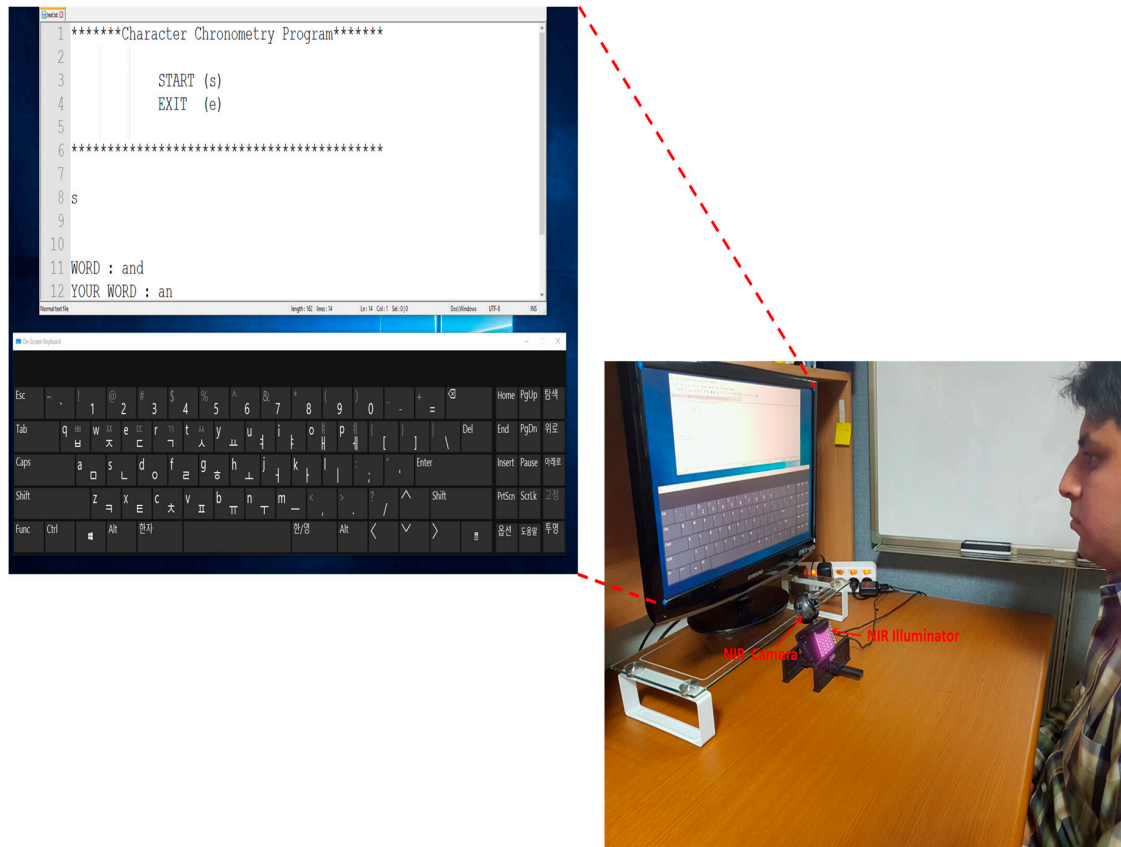

Figure 25. Example of experiment where a user types words on our system.

Table 11. Twenty sample words used for experiments.

the, and, that, have, for,
not, with, you, this, but,
his, from, they, say, her,
she, will, one, all, would

We performed a *t*-test to prove that our method is statistically better than the conventional dwell time-based method [13,14], as shown in Figures 26 and 27. In order to immunize our results against participant learning and physiological state, such as fatigue, we gave a 10-min rest to each participant between experiments. Based on [13,14], dwell time for target selection was set at 500 ms. In all cases, our gaze detection method was used for fair comparison, and selection was performed using our method or the dwell time-based method. We conducted our statistical analysis by using four performance criteria, i.e., accuracy, execution time, interest, and convenience.

As shown in Figure 26a, the *t*-test was performed using two samples independent of each other: the user's typing accuracy using our system ($\mu = 89.7$, $\sigma = 7.4$) and the conventional dwell time-based method ($\mu = 67.5$, $\sigma = 6.5$). The calculated *p*-value was approximately 1.72×10^{-9}, smaller than for a 99% (0.01) significance level. Hence, the null hypothesis for the *t*-test, i.e., no difference between two independent samples, was rejected. Therefore, we concluded that there was a significant difference of up to 99% in accuracy between the proposed method and the dwell time-based method.
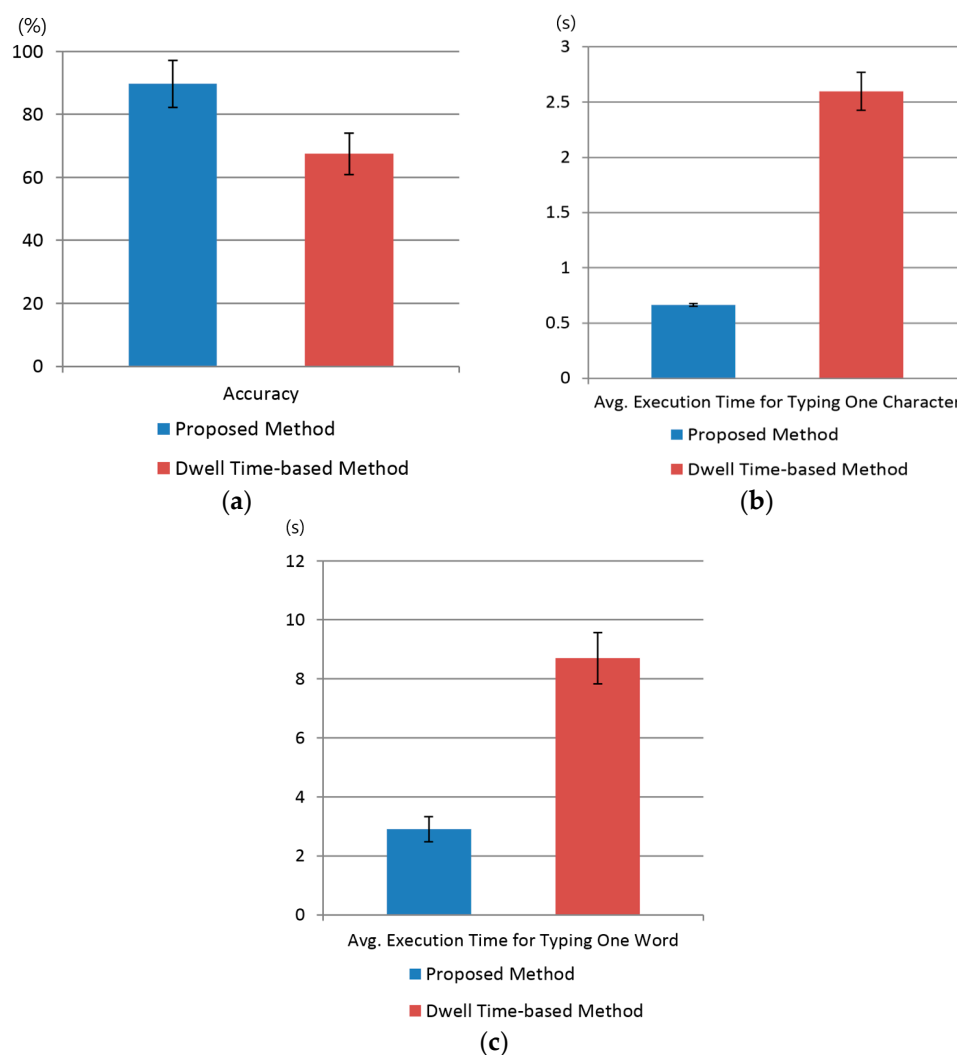

Figure 26. The comparative results of proposed method with dwell time-based method using virtual keyboard in terms of accuracy and execution time. (**a**) Accuracy; (**b**) Average execution time for typing one character; (**c**) Average execution time for typing one word.

Similarly, a *t*-test analysis based on average execution time for typing one character was performed, as shown in Figure 26b, this test yielded similar results, i.e., the average execution time for typing one character with our system ($\mu = 0.67$, $\sigma = 0.013$) and the dwell time-based system ($\mu = 2.6$, $\sigma = 0.17$).

The calculated *p*-value was 2.36×10^{-16}, i.e., smaller than the significance level of 99% (0.01). Therefore, we concluded that there was a significant difference in execution time for typing one character on a virtual keyboard between our system and the dwell time-based system. Although dwell time for target selection was set at 500 ms in the dwell-time-based method, it was often case that the participant did not wait for this long and move his or her gaze position, which increased the average execution time for typing one character, as shown in Figure 26b.

A *t*-test analysis based on execution time for typing one word was also conducted as shown in Figure 26c. We compared our proposed system ($\mu = 2.9$, $\sigma = 0.42$) with the dwell time-based system ($\mu = 8.7$, $\sigma = 0.87$). The calculated *p*-value was 5.7×10^{-16}, i.e., smaller than the significance level of 99% (0.01). Therefore, we concluded that there was a significant difference between the execution time for typing one word on a virtual keyboard between our system and the dwell time-based system.

As shown in Figure 27, the *t*-test analysis for user interest compared our proposed system ($\mu = 4.0$, $\sigma = 0.4$) with the dwell time-based system ($\mu = 2.4$, $\sigma = 0.9$). The calculated *p*-value was 6.1×10^{-7}, i.e., smaller than the significance level of 99% (0.01). Therefore, we concluded that there was a significant difference in user interest between our system and the dwell time-based system.
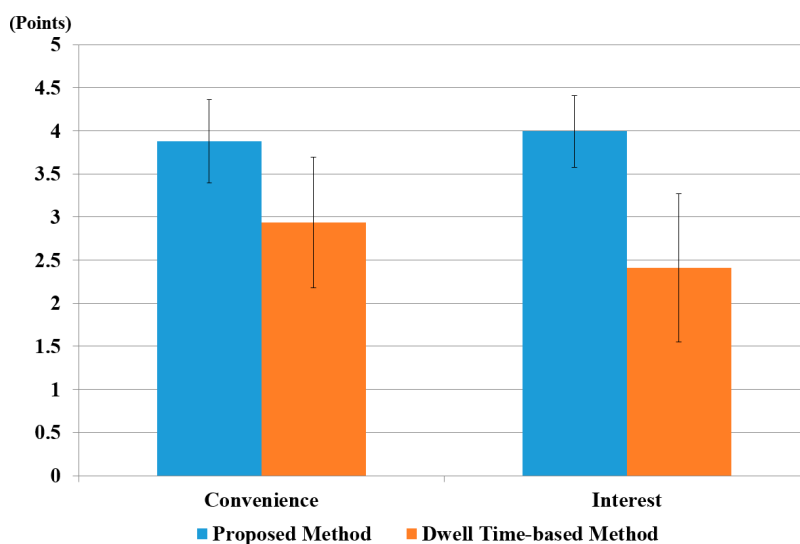

Figure 27. The comparative results of proposed method, with the dwell time-based method using virtual keyboard in terms of convenience and interest.

Similarly, we performed a *t*-test with respect to user convenience. We noted that with our system ($\mu = 3.9$, $\sigma = 0.5$) and the conventional dwell time-based method ($\mu = 2.9$, $\sigma = 0.8$), the calculated *p*-value was approximately 3.4×10^{-4}, which was smaller than the 99% (0.01) significance level. Hence, the null hypothesis for the *t*-test, i.e., no difference between two independent samples was rejected. Therefore, we concluded a significant difference up to 99% in user convenience with our proposed method and the dwell time-based method. As shown in Figure 27, the average score for convenience and interest was higher for the proposed method because it is more natural than a conventional dwell time-based method.

Similarly, we calculated the value of Cohen's *d* for convenience, interest, average execution time for typing, and accuracy. For user convenience and interest, they were 1.49 and 2.34, respectively, and lay in a large category. Hence, this had a significant effect on the difference between the two groups. For average execution times for one character, one word, and accuracy, we calculated the value of Cohen's *d* as approximately 15.92, 8.51, and 3.19, respectively, which also lay in the large-effect category. Hence, from the *p* value and Cohen's *d*, we concluded that user convenience, interest, average execution time, and accuracy were significantly different for the proposed and the dwell time-based methods [13,14].

In our experiment, the screen resolution is 1680×1050 pixels on a 19-inch monitor. The z-distance between the monitor and user's eye ranges 60 to 70 cm. Considering the accuracy (about $\pm 1°$) of gaze detection in our system, the minimum distance between two objects on the monitor screen should be about 2.44 cm (70 cm $\times \tan(2°)$) which corresponds to approximately 82 pixels.

5. Conclusions

In this study, we proposed a method for detecting the user's gaze for target selection using a gaze tracking system based on an NIR illuminator and a camera. The pupil center with the boundary and the glint center were more accurately based on an enhanced Chan–Vese algorithm. We employed three features—change of pupil size w.r.t. time measured by template matching, change in gaze position within short dwell time, and the texture information of monitor image at gaze target. These features as input values were combined with a fuzzy system where optimal input membership functions were designed based on the maximum entropy criterion and the user's gaze for target selection was determined through defuzzification methods. Performance was evaluated by comparing the defuzzification results with the EER and ROC curves. We verified from the results that the COG method using the MIN rule is suitable in terms of accuracy for different objects with varying amounts of texture. In future work, we will investigate a method to enhance performance by combining our proposed features with various physiological ones, such as electrocardiography (ECG) data, electroencephalogram (EEGs) data, or skin temperature (SKT).

Author Contributions: Rizwan Ali Naqvi and Kang Ryoung Park designed the overall system for target selection. In addition, they wrote and revised the paper. Muhammad Arsalan helped designing the fuzzy rule tables, comparative experiments, and collecting databases.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bolt, R.A. Eyes at the Interface. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Gaithersburg, MD, USA, 15–17 March 1982; pp. 360–362.
2. Mauri, C.; Granollers, T.; Lorés, J.; García, M. Computer Vision Interaction for People with Severe Movement Restrictions. *Interdiscip. J. Hum. ICT Environ.* **2006**, *2*, 38–54. [CrossRef]
3. Biosignal. Available online: https://en.wikipedia.org/wiki/Biosignal (accessed on 15 July 2016).
4. Pinheiro, C.G., Jr.; Naves, E.L.M.; Pino, P.; Losson, E.; Andrade, A.O.; Bourhis, G. Alternative Communication Systems for People with Severe Motor Disabilities: A Survey. *Biomed. Eng. Online* **2011**, *10*, 1–28. [CrossRef] [PubMed]
5. Lin, C.-S.; Ho, C.-W.; Chen, W.-C.; Chiu, C.-C.; Yeh, M.-S. Powered Wheelchair Controlled by Eye-Tracking System. *Opt. Appl.* **2006**, *36*, 401–412.
6. Kocejko, T.; Bujnowski, A.; Wtorek, J. Eye Mouse for Disabled. In Proceedings of the Conference on Human System Interactions, Krakow, Poland, 25–27 May 2008; pp. 199–202.
7. Su, M.-C.; Wang, K.-C.; Chen, G.-D. An Eye Tracking System and Its Application in Aids for People with Severe Disabilities. *Biomed. Eng. Appl. Basis Commun.* **2006**, *18*, 319–327. [CrossRef]
8. Heo, H.; Lee, J.M.; Jung, D.; Lee, J.W.; Park, K.R. Nonwearable Gaze Tracking System for Controlling Home Appliances. *Sci. World J.* **2014**, *2014*, 303670. [CrossRef] [PubMed]
9. Ashtiani, B.; MacKenzie, I.S. BlinkWrite2: An Improved Text Entry Method Using Eye Blinks. In Proceedings of the Symposium on Eye-Tracking Research and Applications, Austin, TX, USA, 22–24 March 2010; pp. 339–346.

10. Doughty, M.J. Further Assessment of Gender- and Blink Pattern-Related Differences in the Spontaneous Eyeblink Activity in Primary Gaze in Young Adult Humans. *Optom. Vis. Sci.* **2002**, *79*, 439–447. [CrossRef] [PubMed]

11. Jacob, R.J.K. Eye Movement-Based Human-Computer Interaction Techniques: Toward Non-Command Interfaces. In *Advances in Human-Computer Interaction*; Ablex Publishing Co.: Norwood, NJ, USA, 1993.

12. Majaranta, P.; Räihiä, K.-J. Twenty Years of Eye Typing: Systems and Design Issues. In Proceedings of the Symposium on Eye Tracking Research and Applications, New Orleans, LA, USA, 25–27 March 2002; pp. 15–22.

13. Hansen, J.P.; Tørning, K.; Johansen, A.S.; Itoh, K.; Aoki, H. Gaze Typing Compared With Input by Head and Hand. In Proceedings of the Symposium on Eye Tracking Research and Applications, San Antonio, TX, USA, 22–24 March 2004; pp. 131–138.

14. Hornof, A.J.; Cavender, A. EyeDraw: Enabling Children With Severe Motor Impairments to Draw with Their Eyes. In Proceedings of the Conference on Human Factors in Computing Systems (CHI'05), Portland, OR, USA, 2–7 April 2005; pp. 161–170.

15. Huckauf, A.; Urbina, M.H. Object Selection in Gaze Controlled Systems: What You Don't Look at is What You Get. *ACM Trans. Appl. Percept.* **2011**, *8*, 13:1–13:14. [CrossRef]

16. Ware, C.; Mikaelian, H.H. An Evaluation of an Eye Tracker as a Device for Computer Input. In Proceedings of the SIGCHI/GI Conference on Human Factors in Computing Systems and Graphics Interface, Toronto, ON, Canada, 5–9 April 1987; pp. 183–188.

17. Jacob, R.J.K. The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look at Is What You Get. *ACM Trans. Inform. Syst.* **1991**, *9*, 152–169. [CrossRef]

18. Majaranta, P.; Ahola, U.-K.; Špakov, O. Fast Gaze Typing with an Adjustable Dwell Time. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA, USA, 4–9 April 2009; pp. 357–360.

19. Zhai, S.; Morimoto, C.; Ihde, S. Manual and Gaze Input Cascaded (MAGIC) Pointing. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Pittsburgh, PA, USA, 15–20 May 1999; pp. 246–253.

20. Kumar, M.; Paepcke, A.; Winograd, T. EyePoint: Practical Pointing and Selection Using Gaze and Keyboard. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 28 April–3 May 2007; pp. 421–430.

21. Surakka, V.; Illi, M.; Isokoski, P. Gazing and Frowning as a New Human-Computer Interaction Technique. *ACM Trans. Appl. Percept.* **2004**, *1*, 40–56. [CrossRef]

22. Grauman, K.; Betke, M.; Lombardi, J.; Gips, J.; Bradski, G.R. Communication via Eye Blinks and Eyebrow Raises: Video-Based Human-Computer Interfaces. *Univ. Access Inf. Soc.* **2003**, *2*, 359–373. [CrossRef]

23. Tuisku, O.; Surakka, V.; Rantanen, V.; Vanhala, T.; Lekkala, J. Text Entry by Gazing and Smiling. *Adv. Hum. Comput. Interact.* **2013**, *2013*, 218084. [CrossRef]

24. Naqvi, R.A.; Park, K.R. Discriminating Between Intentional and Unintentional Gaze Fixation Using Multimodal-Based Fuzzy Logic Algorithm for Gaze Tracking System with NIR Camera Sensor. *Opt. Eng.* **2016**, *55*, 063109-1–063109-17. [CrossRef]

25. Momtaz, H.Z.; Daliri, M.R. Predicting the Eye Fixation Locations in the Gray Scale Images in the Visual Scenes with Different Semantic Contents. *Cogn. Neurodyn.* **2016**, *10*, 31–47. [CrossRef] [PubMed]

26. Borji, A.; Sihite, D.N.; Itti, L. Quantitative Analysis of Human-Model Agreement in Visual Saliency Modeling: A Comparative Study. *IEEE Trans. Image Process.* **2013**, *22*, 55–69. [CrossRef] [PubMed]

27. Wang, J.; Borji, A.; Kuo, C.-C.J.; Itti, L. Learning a Combined Model of Visual Saliency for Fixation Prediction. *IEEE Trans. Image Process.* **2016**, *25*, 1566–1579. [CrossRef] [PubMed]

28. Borji, A.; Itti, L. State-of-the-Art in Visual Attention Modeling. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 185–207. [CrossRef] [PubMed]

29. Borji, A.; Tavakoli, H.R.; Sihite, D.N.; Itti, L. Analysis of Scores, Datasets, and Models in Visual Saliency Prediction. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 921–928.

30. Meur, O.L.; Liu, Z. Saccadic Model of Eye Movements for Free-Viewing Condition. *Vis. Res.* **2015**, *116*, 152–164. [CrossRef] [PubMed]

31. Liu, Z.; Shi, R.; Shen, L.; Xue, Y.; Ngan, K.N.; Zhang, Z. Unsupervised Salient Object Segmentation Based on Kernel Density Estimation and Two-Phase Graph Cut. *IEEE Trans. Multimedia* **2012**, *14*, 1275–1289. [CrossRef]

32. Shi, R.; Liu, Z.; Du, H.; Zhang, X.; Shen, L. Region Diversity Maximization for Salient Object Detection. *IEEE Signal Process. Lett.* **2012**, *19*, 215–218. [CrossRef]

33. Guo, C.; Zhang, L. A Novel Multiresolution Spatiotemporal Saliency Detection Model and Its Applications in Image and Video Compression. *IEEE Trans. Image Process.* **2010**, *19*, 185–198. [PubMed]

34. Shen, L.; Liu, Z.; Zhang, Z. A Novel H.264 Rate Control Algorithm with Consideration of Visual Attention. *Multimedia Tools Appl.* **2013**, *63*, 709–727. [CrossRef]

35. Ćulibrk, D.; Mirković, M.; Zlokolica, V.; Pokrić, M.; Crnojević, V.; Kukolj, D. Salient Motion Features for Video Quality Assessment. *IEEE Trans. Image Process.* **2011**, *20*, 948–958. [CrossRef] [PubMed]

36. Liu, Z.; Zhang, X.; Luo, S.; Meur, O.L. Superpixel-Based Spatiotemporal Saliency Detection. *IEEE Trans. Circuits Syst. Video Technol.* **2014**, *24*, 1522–1540. [CrossRef]

37. Treisman, A.M.; Gelade, G. A Feature-Integration Theory of Attention. *Cogn. Psychol.* **1980**, *12*, 97–136. [CrossRef]

38. Koch, C.; Ullman, S. Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry. *Hum. Neurobiol.* **1985**, *4*, 219–227. [PubMed]

39. Itti, L.; Koch, C.; Niebur, E. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 1254–1259. [CrossRef]

40. Bruce, N.D.B.; Tsotsos, J.K. Saliency Based on Information Maximization. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2006; pp. 155–162.

41. Harel, J.; Koch, C.; Perona, P. Graph-Based Visual Saliency. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2006; pp. 545–552.

42. Liu, T.; Sun, J.; Zhang, N.-N.; Tang, X.; Shum, H.-Y. Learning to Detect a Salient Object. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 353–367. [PubMed]

43. Kong, L.; Duan, L.; Yang, W.; Dou, Y. Salient Region Detection: An Integration Approach Based on Image Pyramid and Region Property. *IET Comput. Vis.* **2015**, *9*, 85–97. [CrossRef]

44. Fang, Y.; Lin, W.; Chen, Z.; Tsai, C.-M.; Lin, C.-W. A Video Saliency Detection Model in Compressed Domain. *IEEE Trans. Circuits Syst. Video Technol.* **2014**, *24*, 27–38. [CrossRef]

45. Jiang, H.; Wang, J.; Yuan, Z.; Liu, T.; Zheng, N.; Li, S. Automatic Salient Object Segmentation Based on Context and Shape Prior. In Proceedings of the British Machine Vision Conference, Dundee, UK, 29 August–2 September 2011; pp. 1–12.

46. Cerf, M.; Frady, E.P.; Koch, C. Faces and Text Attract Gaze Independent of the Task: Experimental Data and Computer Model. *J. Vis.* **2009**, *9*, 1–15. [CrossRef] [PubMed]

47. Judd, T.; Ehinger, K.; Durand, F.; Torralba, A. Learning to Predict Where Humans Look. In Proceedings of the International Conference on Computer Vision, Kyoto, Japan, 27 September–4 October 2009; pp. 2106–2113.

48. Chang, K.-Y.; Liu, T.-L.; Chen, H.-T.; Lai, S.-H. Fusing Generic Objectness and Visual Saliency for Salient Object Detection. In Proceedings of International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 914–921.

49. Nuthmann, A.; Henderson, J.M. Object-based Attentional Selection in Scene Viewing. *J. Vis.* **2010**, *10*, 1–19. [CrossRef] [PubMed]

50. Itti, L.; Koch, C. Feature Combination Strategies for Saliency-based Visual Attention Systems. *J. Electron. Imag.* **2001**, *10*, 161–169. [CrossRef]

51. Zhao, Q.; Koch, C. Learning a Saliency Map Using Fixated Locations in Natural Scenes. *J. Vis.* **2011**, *11*, 1–15. [CrossRef] [PubMed]

52. Borji, A. Boosting Bottom-up and Top-down Visual Features for Saliency Estimation. In Proceedings of the International Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 438–445.

53. Reinagel, P.; Zador, A.M. Natural Scene Statistics at the Centre of Gaze. *Netw. Comput. Neural Syst.* **1999**, *10*, 1–10. [CrossRef]

54. Xie, Y.; Lu, H.; Yang, M.-H. Bayesian Saliency via Low and Mid Level Cues. *IEEE Trans. Image Process.* **2013**, *22*, 1689–1698. [PubMed]

55. Li, X.; Lu, H.; Zhang, L.; Ruan, X.; Yang, M.-H. Saliency Detection via Dense and Sparse Reconstruction. In Proceedings of the International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 2976–2983.

56. Borji, A.; Cheng, M.-M.; Jiang, H.; Li, J. Salient Object Detection: A Benchmark. *IEEE Trans. Image Process.* **2015**, *24*, 5706–5722. [CrossRef] [PubMed]

57. Borji, A.; Ahmadabadi, M.N.; Araabi, B.N. Cost-sensitive Learning of Top-Down Modulation for Attentional Control. *Mach. Vis. Appl.* **2011**, *22*, 61–76. [CrossRef]

58. Webcam C600. Available online: http://www.logitech.com/en-us/support/webcams/5869 (accessed on 28 June 2016).

59. Ji, Q.; Yang, X. Real-Time Eye, Gaze, and Face Pose Tracking for Monitoring Driver Vigilance. *Real-Time Imaging* **2002**, *8*, 357–377. [CrossRef]

60. He, Y. Key Techniques and Methods for Imaging Iris in Focus. In Proceedings of the IEEE International Conference on Pattern Recognition, Hong Kong, China, 20–24 August 2006; pp. 557–561.

61. LC Technologies, Inc. Available online: http://www.eyegaze.com/ (accessed on 29 June 2016).

62. SMI. Available online: http://www.smivision.com/ (accessed on 29 June 2016).

63. TheEyeTribe. Available online: https://theeyetribe.com/ (accessed on 29 June 2016).

64. Tobii. Available online: http://www.tobii.com/ (accessed on 29 June 2016).

65. Circular Hough Transform. Available online: https://en.wikipedia.org/wiki/Circle_Hough_Transform (accessed on 14 December 2016).

66. Chan, T.F.; Vese, L.A. Active Contours without Edges. *IEEE Trans. Image Process.* **2001**, *10*, 266–277. [CrossRef] [PubMed]

67. Kass, M.; Witkin, A.; Terzopoulos, D. Snakes: Active Contour Models. *Int. J. Comput. Vis.* **1988**, *1*, 321–331. [CrossRef]

68. Daugman, J. New Methods in Iris Recognition. *IEEE Trans. Syst. Man Cybern. Part B-Cybern.* **2007**, *37*, 1167–1175. [CrossRef]

69. Verney, S.P.; Granholm, E.; Marshall, S.P. Pupillary Responses on the Visual Backward Masking Task Reflect General Cognitive Ability. *Int. J. Psychophysiol.* **2004**, *52*, 23–36. [CrossRef] [PubMed]

70. Hess, E.H. *The Tell-Tale Eye: How Your Eyes Reveal Hidden Thoughts and Emotions*; Van Nostrand Reinhold Co.: New York, NY, USA, 1975.

71. Baddeley, R.J.; Tatler, B.W. High frequency edges (but not contrast) predict where we fixate: A Bayesian system identification analysis. *Vision Res.* **2006**, *46*, 2824–2833. [CrossRef] [PubMed]

72. Yanulevskaya, V.; Marsman, J.B.; Cornelissen, F.; Geusebroek, J.-M. An Image Statistics–Based Model for Fixation Prediction. *Cogn. Comput.* **2011**, *3*, 94–104. [CrossRef] [PubMed]

73. Gabor filter. Available online: https://en.wikipedia.org/wiki/Gabor_filter (accessed on 12 August 2016).

74. Cheng, H.D.; Chen, J.R.; Li, J. Threshold Selection Based on Fuzzy C-partition Entropy Approach. *Pattern Recogn.* **1998**, *31*, 857–870. [CrossRef]

75. Bayu, B.S.; Miura, J. Fuzzy-based Illumination Normalization for Face Recognition. In Proceedings of the IEEE Workshop on Advanced Robotics and Its Social Impacts, Tokyo, Japan, 7–9 November 2013; pp. 131–136.

76. Barua, A.; Mudunuri, L.S.; Kosheleva, O. Why Trapezoidal and Triangular Membership Functions Work So Well: Towards a Theoretical Explanation. *J. Uncertain Syst.* **2014**, *8*, 164–168.

77. Zhao, J.; Bose, B.K. Evaluation of Membership Functions for Fuzzy Logic Controlled Induction Motor Drive. In Proceedings of the IEEE Annual Conference of the Industrial Electronics Society, Sevilla, Spain, 5–8 November 2002; pp. 229–234.

78. Leekwijck, W.V.; Kerre, E.E. Defuzzification: Criteria and Classification. *Fuzzy Sets Syst.* **1999**, *108*, 159–178. [CrossRef]

79. Broekhoven, E.V.; Baets, B.D. Fast and Accurate Center of Gravity Defuzzification of Fuzzy System Outputs Defined on Trapezoidal Fuzzy Partitions. *Fuzzy Sets Syst.* **2006**, *157*, 904–918. [CrossRef]

80. Student's *T*-Test. Available online: http://en.wikipedia.org/wiki/Student\T1\textquoterights_t-test (accessed on 5 July 2016).

81. Cohen, J. A Power Primer. *Psychol. Bull.* **1992**, *112*, 155–159. [CrossRef] [PubMed]

82. Most Common Words in English. Available online: http://en.wikipedia.org/wiki/Most_common_words_in_English (accessed on 19 March 2014).