

DOCTORAL DISSERTATION

HIGH PRECISION MOVING OBJECT DETECTION FOR
VIDEO OCULOGRAPHY AND TRAFFIC MONITORING

(ビデオ眼球運動および交通監視のための高精度移動物体検出)

SEPTEMBER, 2021

YOANDA ALIM SYAHBANA

GIFU UNIVERSITY

DOCTORAL DISSERTATION

HIGH PRECISION MOVING OBJECT DETECTION FOR VIDEO OCULOGRAPHY AND TRAFFIC MONITORING

SEPTEMBER, 2021



Electronics and Information Systems Engineering Division

Graduate School of Engineering

Gifu University

Japan

YOANDA ALIM SYAHBANA

HIGH PRECISION MOVING OBJECT DETECTION FOR VIDEO OCULOGRAPHY AND TRAFFIC MONITORING

by

YOANDA ALIM SYAHBANA

**Submitted for the degree of
Doctor of Philosophy in Engineering**



**Yokota Laboratory
Electronics and Information Systems Engineering Division
Graduate School of Engineering
Gifu University
Japan**

September, 2021

ABSTRACT

Generally, the research objective of moving object detection is to observe persons or objects on the move and to provide information of objects position in a timely ordered sequence. The use of video cameras to achieve the objective has become popular with camera and computing resource advancement. This study takes two study cases of moving object detection as a theme of research.

As a first study case, it is focused on the topic of moving object detection from video oculography that is used to detect nystagmus from patients. The detection, in this case, has accuracy problems when patients who complain of dizziness have difficulty fully opening their eyes. Pupil detection and tracking in this condition affect the accuracy of the nystagmus waveform. This study designs a pupil detection method using pattern matching that approximates the pupil. In order to estimate patient pupil position accurately, it is practical to model and use pupil shape. In general, the current study uses circle shape to approximate pupil shape, such as the circular Hough transform method. However, the actual pupil shape is slightly flattened from a perfect circle and forms an ellipse. Approximating ellipse shape with the model of circle causes deterioration of pupil estimation accuracy. Therefore, this study uses a Mexican-hat-type ellipse pattern to deal with the problem. This study evaluates the proposed method performance to conventional Hough transform method for 37 eye movement videos retrieved from Gifu University Hospital. This study uses Infrared Eye Movement Imaging TV Device IEM-2 from Nagashima Medical Instrument Co. Ltd. as eye movement observation equipment. In this study, approximating pupil using ellipse shape increases parameter to be estimated and calculation cost while compared to a circle shape. Therefore, this study adopts a method of improving estimation accuracy in three steps: rough, precise, and subpixel detection to estimate pupil centroid and radius. The proposed method performance is evaluated for the eye movement video because the video only shows a partial part of the pupil. Performance results show that the proposed method can detect and track pupil position even though only 20% of the pupil is visible. In comparison, the conventional Hough transform only indicates good performance if 90% of the pupil is visible. This study also evaluates the proposed method using the Labelled Pupil in the Wild (LPW) dataset. This data set has been labeled with pupil centroid information as ground truth for performance evaluation. From the LPW dataset, this study selects a total of 675 eye images. Selection of pupil images is conducted based on pupil images of the respondent that does not use glasses, eye contacts, and mascara in an indoor situation without strong reflection. This

study also selects pupil images captured from the front side to resemble typical nystagmus observation images. The result shows that the proposed method accuracy has 1.47 of Mean Square Error (MSE), lower than the conventional Hough transform method with 9.53 of MSE. This study conducts expert validation by consulting the nystagmus waveform with three medical specialists. The medical specialists agree that the waveform can be evaluated clinically without contradiction to their diagnosis.

As the second study case, it is focused on traffic monitoring video deployed for adaptive traffic light during temporary roadblock that only allows one side traffic flow at a time. It is becoming necessary to detect and track incoming traffic while only one side of road lane is accessible at one time during temporary road block. As such, it is crucial that distant incoming traffic be recognized as early as possible, as waiting for the incoming traffic too close to the traffic light could be too late for signaling and lead to sudden breaking or may even cause an accident. The purpose of this study was to develop improved detection and tracking the traffic even as the traffic are still distant from the traffic light. This study formulates the proposed method as three steps. The proposed method is initialization step that consist of two processes. The first process is to enclose region where the main movement of traffic exist in captured scene. In this process, this study uses background subtraction method to detect foreground object followed by frame difference method to detect movement of the object. Then, the object movement is aggregated to define RoI. The second process is aimed to estimate coordinate of vanishing point from the video frame. This study uses WOD method that calculates differential excitation of the frame texture features. Then, Gabor filter is used to calculate pixel orientation. The information of differential excitation and orientation from every pixel is used for voting scheme to estimate vanishing point coordinate. The second step is aimed to focus the detection of traffic within the RoI. Similar foreground object detection in the initialization step is used to detect the traffic. Following this, the third step is aimed track the detected traffic. This step associates the detected traffic based on its movement from frame to frame. Kalman filter is used track the detected traffic based on likelihood of each detection to each motion track. In addition, the motion track is selected for the traffic that getting distant from the vanishing point. Finally, the selected motion track and its detected traffic are categorized as the incoming traffic. Evaluation is conducted based on how early the proposed method in detecting incoming traffic compared to R-CNN method. The proposed method requires average of 17.75 frames to detect the target vehicle while the R-CNN requires average of 63.36 frames to detect the target vehicle. The finding suggest that the accuracy of proposed method is depended on number of pixel orientation in

estimating the vanishing point and definition of RoI. Therefore, the proposed method is reliable to support safety and reliability of adaptive traffic light system.

ACKNOWLEDGEMENTS

This dissertation could not have been fulfilled if it were not for the understanding and support of the following people:

My deepest gratitude goes first and foremost to my supervisor Prof. Yasunari Yokota for guidance and meaningful ideas during the research discussion. In addition, He also provides tireless effort and teaching that motivate me to be a better researcher and lecturer during my three-years study at Gifu University.

I also would like to express my gratitude to Prof. Takeshi Hara and Associate Prof. Xiangrong Zhou for their valuable discussion through the dissertation defense.

I also thank all the members of Yokota laboratory for the daily discussion on the related matter.

Finally, I also would like to express my deep gratitude to those who have contributed in one way or another to the completion of this work and doctoral degree acquisition. Thank you to my wife, parent, family, AGP program from Gifu University, and Politeknik Caltex Riau. Last but not least, I would like to thank you for your interest in my dissertation.

CONTENTS

ABSTRACT	V
ACKNOWLEDGEMENTS.....	VIII
1 INTRODUCTION.....	1
1.1 NYSTAGMUS ESTIMATION FOR DIZZINESS DIAGNOSIS BY PUPIL DETECTION AND TRACKING USING MEXICAN-HAT-TYPE ELLIPSE PATTERN MATCHING.....	1
1.2 EARLY DETECTION AND TRACKING OF DISTANT INCOMING TRAFFIC USING IMPROVED DETECTION ON ROAD VANISHING POINT REFERENCE FOR ADAPTIVE TRAFFIC LIGHT SIGNALING	3
2 NYSTAGMUS ESTIMATION FOR DIZZINESS DIAGNOSIS BY PUPIL DETECTION AND TRACKING USING MEXICAN-HAT-TYPE ELLIPSE PATTERN MATCHING.....	5
2.1 WORKING PRINCIPLE OF THE EYE MOVEMENT OBSERVATION EQUIPMENT	5
2.2 DATA SET DESCRIPTION.....	6
2.2.1 <i>Eye Movement Video from Gifu University Hospital</i>	6
2.2.2 <i>Labelled Pupil in the Wild (LPW) Data Set</i>	7
2.3 PROPOSED METHOD	7
2.3.1 <i>Infrared Spot Filling</i>	8
2.3.2 <i>Mexican Hat-Type Ellipse Pattern Matching</i>	9
2.3.3 <i>Three Steps Precision Improvement</i>	11
2.3.4 <i>Estimation of the Optimal Flatness Parameter q</i>	12
2.4 RESULTS.....	13
2.5 EVALUATION	15
2.5.1 <i>Performance Evaluation for Partially Shown Pupil</i>	15
2.5.2 <i>Performance Evaluation for Partially Shown Pupil</i>	19
2.5.3 <i>Medical Specialist Validation</i>	20
2.6 CONCLUSION	23
3 EARLY DETECTION AND TRACKING OF DISTANT INCOMING TRAFFIC USING IMPROVED DETECTION ON ROAD VANISHING POINT REFERENCE FOR ADAPTIVE TRAFFIC LIGHT SIGNALING.....	25
3.1 MATERIALS AND METHODS.....	25
3.1.1 <i>Video Test Material</i>	25

3.1.2 <i>Proposed Method</i>	26
3.2 RESULT AND DISCUSSION.....	30
3.2.1 <i>Initialization Step</i>	30
3.2.2 <i>Detection and Tracking of Incoming Traffic</i>	34
3.3 EVALUATION.....	35
3.4 CONCLUSIONS	37
LIST OF PUBLICATIONS	38
REFERENCES.....	39
APPENDICES	43
APPENDIX A: SUMMARY OF SUBJECT VIDEO AND MEDICAL SPECIALIST REVIEW.....	44
APPENDIX B: CALCULATION METHOD FOR THE MAGNITUDE OF FLUCTUATION	50

LIST OF TABLES

TABLE 3.1 QUANTITATIVE COMPARISON OF VANISHING POINT ESTIMATION ERROR (δ) FOR VARIATION OF $N\phi$	34
TABLE 3.2 BENCHMARKING RESULT ON DETECTION OF INCOMING TRAFFIC.....	37

LIST OF FIGURES

FIGURE 2.1: EYE MOVEMENT OBSERVATION EQUIPMENT: (A) INFRARED EYE MOVEMENT IMAGING TV DEVICE IEM-2 AND VIDEO CAPTURE; AND (B) SYSTEM ILLUSTRATION.	5
FIGURE 2.2: ILLUSTRATION OF EYE MOVEMENT OBSERVATION EQUIPMENT.....	6
FIGURE 2.3: DESIGN OF PROPOSED METHOD.....	8
FIGURE 2.4: THE EXAMPLE OF THE FUNCTION $f(x, y; x_0, y_0, r, q)$, WITH $x_0 = y_0 = 0$, $r = 8$, AND $q = 0.90$: (A) BIRD'S-EYE VIEW; AND (B) CROSS-SECTION AT $Y = 0$.	10
FIGURE 2.5: SAMPLE OF $r(t)$ FOR VARYING VALUES OF q	13
FIGURE 2.6: SAMPLE OF INFRARED SPOT FILLING TO THE DETECTION RESULT.....	13
FIGURE 2.7: COMPARISON OF ROUGH, PRECISE, AND SUBPIXEL DETECTION RESULTS.....	14
FIGURE 2.8: SAMPLE OF THE NYSTAGMUS WAVEFORM GENERATED BY THE PROPOSED METHOD.	15
FIGURE 2.9: ILLUSTRATION OF PUPIL CROPPING: (A) 100%; (B) 90%; (C) 80%; (D) 70%; (E) 60%; (F) 50%; (G) 40%; (H) 30%; (I) 20%; AND (J) 10%.....	15
FIGURE 2.10: COMPARISON OF MSE CALCULATION RESULTS FROM THE PROPOSED MEXICAN HAT-TYPE ELLIPSE PATTERN MATCHING AND THE CONVENTIONAL HOUGH TRANSFORM METHOD.	17
FIGURE 2.11: THE EXAMPLE OF THE CONVENTIONAL HOUGH TRANSFORM PATTERN WITH A UNIFORM-VALUED RING: (A) BIRD'S-EYE VIEW; AND (B) CROSS-SECTION AT $Y = 0$..	18
FIGURE 2.12: THE DIFFERENCE IN PEAK SHARPNESS FOR THE EVALUATION FUNCTION $h(x_0, y_0, r, q; t)$: (A) CONVENTIONAL HOUGH TRANSFORM; AND (B) MEXICAN HAT-TYPE ELLIPSE PATTERN.	19
FIGURE 2.13: NYSTAGMUS WAVEFORM FROM THE PROPOSED METHOD FOR VIDEO No. 1, HORIZONTAL MOVEMENT OF THE PUPIL.....	20
FIGURE 2.14: NYSTAGMUS WAVEFORM FROM THE PROPOSED METHOD FOR VIDEO No. 1, VERTICAL MOVEMENT OF THE PUPIL.	21
FIGURE 2.15: NYSTAGMUS WAVEFORM FROM THE PROPOSED METHOD FOR VIDEO No. 28, HORIZONTAL MOVEMENT OF THE PUPIL.....	21
FIGURE 2.16: SAMPLE OF A VIDEO FRAME FROM VIDEO No. 2.	22

FIGURE 2.17: NYSTAGMUS WAVEFORM FOR VIDEO No. 2: (A) USING THE PROPOSED METHOD AND; (B) USING THE CONVENTIONAL HOUGH TRANSFORM METHOD.	22
FIGURE 2.18: SAMPLE OF A VIDEO FRAME FROM VIDEO No. 11.	23
FIGURE 2.19: NYSTAGMUS WAVEFORM FROM THE PROPOSED METHOD FOR VIDEO No. 11, HORIZONTAL MOVEMENT OF THE PUPIL.	23
FIGURE 3.1: SAMPLE OF CAPTURED SCENE FROM VIDEO TEST MATERIAL: (A) SITE 1; (B) SITE 2; (C) SITE 3; (D) SITE 4; (E) SITE 5; (F) SITE 6; (G) SITE 7; (H) SITE 8; (I) SITE 9; (J) SITE 10; (K) SITE 11; (L) SITE 12.	25
FIGURE 3.2: DESIGN OF PROPOSED METHOD.	26
FIGURE 3.3: RESULT OF INITIALIZATION STEP OF THE PROPOSED METHOD: (A) SITE 1; (B) SITE 2; (C) SITE 3; (D) SITE 4; (E) SITE 5; (F) SITE 6; (G) SITE 7; (H) SITE 8; (I) SITE 9; (J) SITE 10; (K) SITE 11; (L) SITE 12.	32
FIGURE 3.4: SAMPLE DETECTED EDGE (FIRST ROW) AND ACCUMULATOR SPACE (SECOND ROW) THAT INFLUENCE THE ESTIMATED VANISHING POINT: (A) SITE 6; (B) SITE 7; (C) SITE 11; (D) SITE 12.	33
FIGURE 3.5: SAMPLE DETECTION OF $IFO_{x,y,t}$ (FIRST ROW) AND FINAL RESULT OF INCOMING TRAFFIC DETECTION (SECOND ROW): (A) SITE 2; (B) SITE 4; (C) SITE 7; (D) SITE 12.	35

LIST OF APPENDICES

APPENDIX A: SUMMARY OF SUBJECT VIDEO AND MEDICAL SPECIALIST REVIEW	44
APPENDIX B: CALCULATION METHOD FOR THE MAGNITUDE OF FLUCTUATION	50

1 INTRODUCTION

Moving object detection covers variety of uses. Some of them are human-computer interaction, surveillance-safety-security system, traffic control and monitoring, and medical imaging [1]. Generally, the objective is to observe a persons or objects on the move then to provide information of object position in timely ordered sequence. The use of camera to achieve the objective has become popular with support of camera and computing resource advancement. In addition, research on computer vision based technique also has been widely studied since 90s and thousands of algorithms has been proposed to achieve the objective with its each challenges.

This study focuses topic of moving object detection in two study cases. The first study case is eye movement video from patient who complain of dizziness. Objective of this study is to detect and track pupil centroid in the video. Then, information of the centroid is plotted as nystagmus waveform that used for medical specialist diagnosis. Challenge in this study case is to achieve the objective in a condition if the pupil is not fully visible due to difficulty of the patient to consciously open their eyes.

The Second study case is traffic monitoring video deployed for adaptive traffic light during temporary roadblock that only allows one side traffic flow at a time. Detection of incoming traffic is influenced by perspective projection that influence early detection of incoming traffic to signal appropriate light.

In order to achieve the objective on the two study cases, this study proposed two image processing methods that are Mexican-hat-type ellipse pattern matching for pupil detection for the first study case, and improved detection on road vanishing point reference for the second study case.

1.1 Nystagmus estimation for dizziness diagnosis by pupil detection and tracking using Mexican-hat-type ellipse pattern matching

Dizziness is a common symptom presented by patients in a health examination [2]. Dizziness represents an unsteady sensation accompanied by a feeling of movement within the head [3]. Based on [4], the four categories of dizziness are lightheadedness, presyncope, disequilibrium, and vertigo. Among these categories, vertigo is the most common cause of dizziness, which is related to neurological conditions [5]. Two categories of vertigo are central vertigo, related to disease/injury in the brain, and

peripheral vertigo, related to a vestibular disorder. In terms of signs and symptoms, vertigo has many potential causes, and the symptoms can be vague, non-specific, and inconsistent [6]. As dizziness due to vertigo is a subjective symptom, the symptom threshold depends on the patient's sensitivity [7]. Correlating dizziness and its cause has become a significant challenge for medical specialists, as the cause of dizziness determines the treatment offered for dizziness [4].

Based on [8]–[11], medical specialists can use nystagmus symptoms as a crucial element in identifying the cause of dizziness. Different types of nystagmus can be categorized by analyzing the fast and slow phase or the alternating slow phase of eye movement. Existing studies of nystagmus have provided a review of the critical clinical literature, in order to support state-of-the-art differential diagnosis [12], and have discussed the clinical features of nystagmus and its relation to ocular motility disorder [13]–[15]. Existing studies have also focused on the treatment and therapy process [16], [17], and case-by-case of nystagmus in specific subject categories [18], [19] for different forms of nystagmus: vertical, positional, head-shaking and vibration-induced, and vestibular nystagmus.

Conventional observation, which is conducted visually by the medical specialist, can be subjectively biased. The visual examination also requires a medical specialist's sufficient experience for accurate diagnosis. Furthermore, patients with dizziness may feel pain when attempting to consciously fully open their eyes, as such, their eye may remain only partially open. Therefore, an emphasis on nystagmus observation to support clinical decisions is essential, in order to enhance diagnostic reasoning by medical specialists [20]. A practical method is required to objectively measure eye movements and present the movement as a nystagmus waveform to the medical specialist.

An alternative method for eye movement measurement is video-oculography [21]–[24]. This method uses a camera to capture eye images, a computer to record the captured images, and software to detect and track eye movement. Due to advancements in camera technology and computer processing capability, video-oculography has become more popular and can serve as a more reliable method to measure eye movement [25]. In this research, we adopt the video-oculography method to obtain a nystagmus waveform for dizziness diagnosis. The waveform presents estimated eye movement, based on tracked pupil position from the patient's eye. Initially, Frenzel goggles (equipped with an infrared camera and infrared illumination) capture eye images under night vision, with the light blocked by the goggles. Similarly, in this research, we use infrared light as a light source

and an infrared camera to capture eye images. The light enters the pupil and diffuses inside the eyeball. Then, tissues and vitreous humor inside the eyeball absorb the diffused light. This process causes the pupil to become dark in the video frame. On the contrary, the iris and sclera will reflect the light, causing these areas to become bright [26]. Therefore, we tracked pupil position based on the high contrast between the iris and pupil, creating a boundary between the dark pupil area and the bright iris area.

In order to estimate patient pupil position accurately under the previously mentioned conditions, it is practical to model and use the pupil shape. In general, existing research uses a circle shape to approximate pupil shape [27], such as the circular Hough transform [24] method; however, the actual pupil shape is slightly flattened from a perfect circle and forms an ellipse. Approximating an ellipse shape with a model based on a circle causes deterioration of pupil estimation accuracy. In addition, a patient who complains of dizziness often has difficulty in opening their eyes fully. Therefore, this research proposes a pupil detection and tracking method using a Mexican-hat-type ellipse pattern, which can detect pupil position for a partially open pupil, as the main contribution of this research.

The result shows that the proposed Mexican hat-type ellipse pattern matching can detect pupil position even if the pupil is opened partially. Compared to Hough transform method, the result also shows that the proposed Mexican hat-type ellipse pattern matching achieves lower Mean Square Error (MSE) result.

1.2 Early Detection and Tracking of Distant Incoming Traffic using Improved Detection on Road Vanishing Point Reference for Adaptive Traffic Light Signaling

Temporary roadblock that only allows one side traffic flow at a time requires traffic controller to manage the traffic. However, the traffic controller is vulnerable to traffic accidents. As an alternative, timing based traffic light can be positioned to control the traffic. Nevertheless, timing based traffic light is not effective because it is not depended on real traffic condition. Therefore, an adaptive traffic light based on real time traffic condition is recommended [28], [29]. The adaptive traffic light is connected to camera over computer to capture real time traffic condition. Decision to signal the lights depends on the traffic condition from both side of road that is captured from the camera.

Early detection of incoming traffic for adaptive traffic light is important challenges. The early detection provides information for decision process of traffic light system to signal

appropriate light. Late decision leads to late signaling that cause the incoming traffic to stop suddenly and may cause an accident. However, captured scene from the camera is influenced by perspective projection. Far incoming traffic appears in small size. Detection this small object is challenging because existence of non-vehicle object. In addition, utilizing artificial intelligence such as deep learning to detect this small object deals with limitation because the vehicle image has low resolution [30]–[32].

This study formulates the proposed method as three steps. The proposed method is initialization step that consist of two processes. The first process is to enclose region where the main movement of traffic exist in captured scene. In this process, this study uses background subtraction method to detect foreground object followed by frame difference method to detect movement of the object. Then, the object movement is aggregated to define RoI. The second process is aimed to estimate coordinate of vanishing point from the video frame. This study uses WOD method [33] that calculates differential excitation of the frame texture features. Then, Gabor filter is used to calculate pixel orientation. The information of differential excitation and orientation from every pixel is used for voting scheme to estimate vanishing point coordinate.

The second step is aimed to focus the detection of traffic within the RoI. Similar foreground object detection in the initialization step is used to detect the traffic. Following this, the third step is aimed track the detected traffic. This step associates the detected traffic based on its movement from frame to frame. Kalman filter is used track the detected traffic based on likelihood of each detection to each motion track. In addition, the motion track is selected for the traffic that getting distant from the vanishing point. Finally, the selected motion track and its detected traffic are categorized as the incoming traffic.

2 NYSTAGMUS ESTIMATION FOR DIZZINESS DIAGNOSIS BY PUPIL DETECTION AND TRACKING USING MEXICAN-HAT-TYPE ELLIPSE PATTERN MATCHING

2.1 Working Principle of the Eye Movement Observation Equipment

Generally, eye movement observations associated with dizziness are conducted by preventing the visual fixation of patients' eyes [4], [13]. Therefore, the observation of nystagmus was conducted under night vision. We used the Infrared Eye Movement Imaging TV Device IEM-2 from Nagashima Medical Instrument Co. Ltd., shown in Figure 2.1a. The device includes wearable goggles with an infrared camera connected to a video decoder. The patient's eyes were positioned inside the goggles to block the light from outside by a cover made of rubber. The goggles are also attached to an infrared light source that illuminates either the left or right eye of the patient; thus, the infrared camera can capture the eye. Then, the TV monitor presents the images captured by the camera, through a computer equipped with a video capture card. Figure 2.1 illustrates the system of eye movement observation equipment.

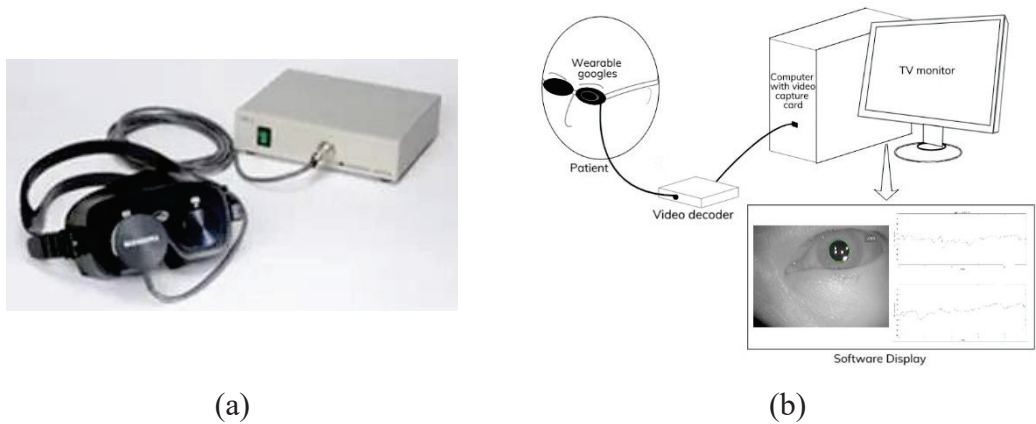


Figure 2.1: Eye movement observation equipment: (a) Infrared Eye Movement Imaging TV Device IEM-2 and video capture; and (b) system illustration.

Eye movement observations using the abovementioned equipment were based on the dark-pupil technique. In this technique, the equipment illuminates the eye with an 887 nm near-infrared (NIR) light source and records the eye image with an infrared camera. The dark-pupil technique causes the pupil to become the darkest region in the image, as the eye is illuminated by an off-axis source. The light enters the pupil and diffuses inside of

the eyeball. Then, the tissues and vitreous humor inside the eyeball absorb the diffused light. On the contrary, the iris, sclera, and eyelids reflect the light and appear bright in the eye image. This research uses an intensity gradient between the pupil and the iris to detect the pupil contour. The light also generates a corneal reflection of the light source, appearing as small and sharp glint dots. From now on, the dots are referred to as infrared spots. Figure 2.2 shows the working principle of the eye movement observation equipment used in this research.

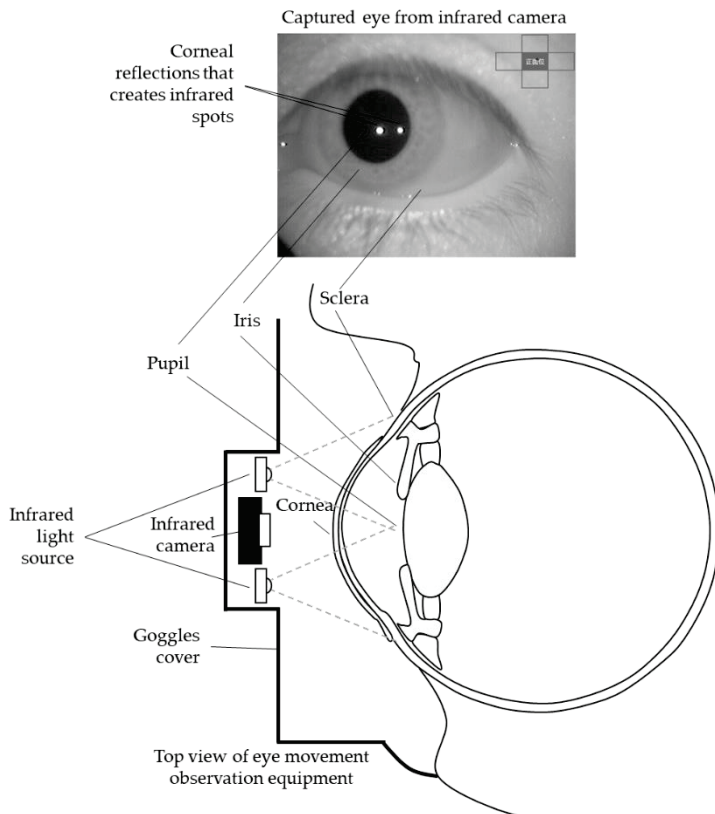


Figure 2.2: Illustration of eye movement observation equipment.

2.2 Data Set Description

For this research, we used two data sets. The primary data set comprises eye movement videos obtained using the eye movement observation equipment explained in Section 2.1. The additional data set is the publicly available Labelled Pupil in the Wild (LPW) data set.

2.2.1 Eye Movement Video from Gifu University Hospital

The subjects in the eye movement videos were 22 males and 15 females aged from 28 to 81 years old. The subjects were diagnosed with semicircular canals or brain-related illnesses, such as Meniere's disease, vestibular disorder, medulla oblongata bleeding,

spinocerebellar degeneration, or multiple system atrophy. The eye movement videos of the subjects were retrieved from Gifu University Hospital. The videos show eye images with regular shape and good pupil transparency conditions. Table A1 in Appendix A summarizes the eye videos from these subjects.

A video frame from an eye movement video can be represented as $I(x, y, t) \in \{0, 1, \dots, 255\}$, $x \in \{1, 2, \dots, N_x\}$, $y \in \{1, 2, \dots, N_y\}$, and $t \in \{1, 2, \dots, T\}$, where N_x and N_y are the width and height of the video frame, respectively, and T is the total number of video frames. The total video frames, T , was calculated as:

$$T = Vduration * Vfps, \quad (2.1)$$

where $Vduration(s)$ is the duration of the video and $Vfps$ (frame/s) is the video's frame rate. In this research, $N_x = 640$ pixels and $N_y = 480$ pixels, except for videos 17, 18, 19, 24, and 27, which had $N_x = 720$ pixels and $N_y = 480$ pixels. In addition, video number 37 had $N_x = 320$ and $N_y = 240$ pixels. In this research, the video frame rate was $Vfps = 30$ frame/s. The total duration, $Vduration$, for each video used in this research is summarized in Table A1 in Appendix A.

2.2.2 Labelled Pupil in the Wild (LPW) Data Set

We evaluated the performance of the proposed method using the LPW data set [34]. This data set has been labeled with pupil center information as ground truth, for performance evaluation [35]. From the LPW data set, we selected $I(x, y)$ for a total of 675 eye images, with $N_x = 384$ pixels and $N_y = 288$ pixels. The selection of $I(x, y)$ was conducted based on pupil images of respondents that did not use glasses, eye contacts, or mascara in an indoor situation without strong reflection. We also selected $I(x, y)$ captured from the front side, such that they resembled typical nystagmus observation images.

2.3 Proposed Method

Figure 2.3 shows the design of the proposed method, which is divided into nine processes. The details of each process are discussed in the following subsections.

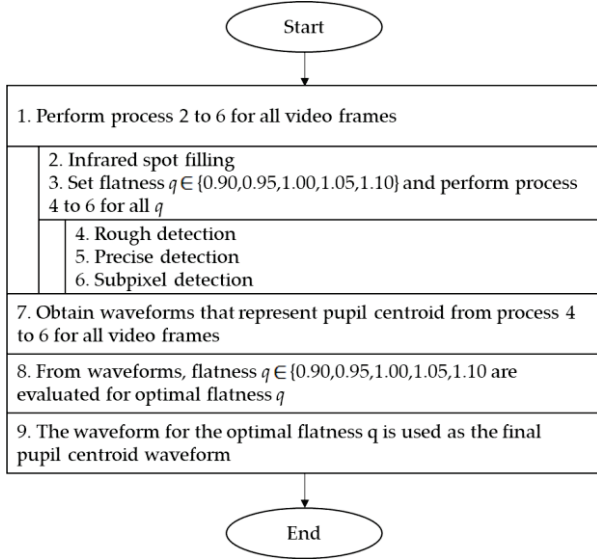


Figure 2.3: Design of proposed method.

2.3.1 Infrared Spot Filling

As previously explained in Section 2.1, infrared light was used as a light source. A transparent membrane can reflect infrared light on the surface of the cornea and create infrared spots. Processing is required to remove the reflected infrared spots in the video frame, as they produce strong edges and adversely affect the estimation of pupil position. The brightness of this infrared spot was approximately represented by a high-intensity value (i.e., larger than 250). Therefore, the spot was detected by

$$I_{spot}(x, y, t) = \begin{cases} 1, & I(x, y, t) > 250 \\ 0, & otherwise \end{cases}, \quad (2.2)$$

where $I_{spot}(x, y, t)$ is the detected reflection of the infrared spot. $I_{spot}(x, y, t)$ is a variable that takes a binary value, representing a pixel estimated to be an infrared spot with 1 and all others with 0. Around these spots, there exist regions with lower intensity values (i.e., $I(x, y, t) < 250$), which are also part of the infrared spot reflection. Therefore, a dilation process was applied, in order to include the surrounding region. $I_{spot}(x, y, t)$ is dilated with a size of 7×7 ; thus, the surrounding region is also detected as an infrared spot. Then, a mean value of pixels in $I(x, y, t)$ that surround over one pixel outside the infrared spot replaces the intensity value in the corresponding $I(x, y, t)$ within the infrared spot region. After this step, $I(x, y, t)$ is redefined as a video frame without an infrared spot.

Edge detection is performed on $I(x, y, t)$ for each frame t . Several popular methods, including Sobel, Prewitt, Roberts, and Canny, were compared for the videos tabulated in

Table A1, Appendix A. Comparing these methods, the Canny method had the best performance, and we decided to use the Canny edge detection method for our experiment. The edge detection result from the image $I(x, y, t)$ is represented by $I_{edge}(x, y, t)$.

2.3.2 Mexican Hat-Type Ellipse Pattern Matching

In order to detect the pupil as an ellipse, it is necessary to estimate the parameters of the ellipse, including the x coordinate, y coordinate, radius, flatness, and flattening direction of the center of the pupil. We confirmed that the pupil is flattened only in the vertical direction and stays equal in the horizontal direction, based on an examination of all eye movement videos. Therefore, the flat direction parameter of the ellipse was only focused on the vertical direction. The ellipse with a radius r centered at the coordinate (x_0, y_0) can be represented the set of points (x, y) satisfying the equation

$$\left(\frac{x-x_0}{q}\right)^2 + (y-y_0)^2 = r^2, \quad (2.3)$$

where q is the flatness of the ellipse, which represents the ratio of the horizontal radius to the vertical radius of the ellipse. As an illustration, a perfect circle is obtained when $q=1$, a horizontally long ellipse is obtained when $q>1$, and a vertically long ellipse is obtained when $q<1$. A pattern matching process was performed on the edge image, $I_{edge}(x, y, t)$, using the generated ellipse pattern. The center coordinate (x_0, y_0) , radius r , and flatness q were obtained by maximizing the evaluation function in the pattern matching process.

In order to define the evaluation function, the following two-dimensional function $f(x, y; x_0, y_0, r, q)$, as the ellipse pattern, was calculated using

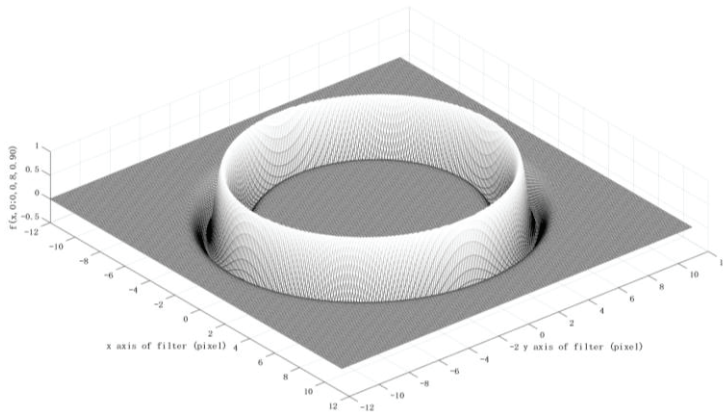
$$f(x, y; x_0, y_0, r, q) = (1 - g(x, y; x_0, y_0, r, q)) e^{\frac{g(x, y; x_0, y_0, r, q)}{2}}, \quad (2.4)$$

in which,

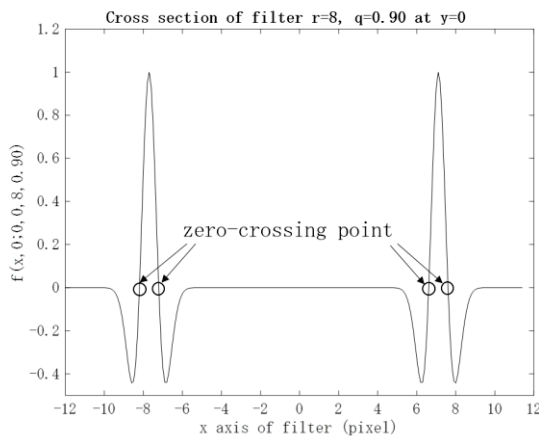
$$g(x, y; x_0, y_0, r, q) = \left(\frac{\sqrt{\left(\frac{x-x_0}{q}\right)^2 + (y-y_0)^2} - r}{\frac{r}{15}} \right)^2. \quad (2.5)$$

An example of the function $f(x, y; x_0, y_0, r, q)$, with $x_0 = y_0 = 0$, $r = 8$, and $q = 0.90$, is shown in Figure 2.4. Figure 2.4a,b shows the bird's-eye view and the cross-section at $y = 0$ of

the function, respectively. The $r/15$ in Equation (2.5) represents the zero-crossing point into lateral suppression, marked by the black circles in Figure 2.4b. This optimal value was determined by some preliminary experiments on all eye movement videos. This Mexican hat-type ellipse pattern aims to concentrate the blurred edge of the pupil into a single sharp peak of the evaluation function. The Mexican hat-type shape will have maximum amplitude at a single peak and gradually suppresses insignificant edges. Therefore, the Mexican hat-type ellipse pattern can improve the accuracy of ellipse detection. A similar approach has also been studied, in order to improve the conventional Hough transform accuracy in detecting circle shapes, instead of the ellipse shape used in this research [36]. The result shows that the Mexican hat-type shape fitted the circle candidate and removed the fake circle associated with the conventional Hough transform. The term Mexican hat is used, due to its similarity to a Sombrero when plotted as a 2D image.



(a)



(b)

Figure 2.4: The example of the function $f(x, y; x_0, y_0, r, q)$, with $x_0 = y_0 = 0$, $r = 8$, and $q = 0.90$: (a) Bird's-eye view; and (b) cross-section at $y = 0$.

Initially, we investigated the ranges of radius and flatness for all eye movement videos for the subjects denoted in Table A1, Appendix A. Based on the investigation results, the radius r and flatness q were approximately varied, as $32 \leq r \leq 104$ pixels and $0.90 \leq q \leq 1.10$, respectively. Thus, the search range of pupil shape was defined, based on the radius r , as $r \in \{32, 36, \dots, 104\}$ and, based on the flatness q , as $q \in \{0.90, 0.95, 1.00, 1.05, 1.10\}$.

The evaluation function, namely, the degree of similarity, was defined as:

$$h(x_0, y_0, r, q; t) = \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} f(x, y; x_0, y_0, r, q) I_{edge}(x, y, t), \quad (2.6)$$

for each frame t and flatness q . The calculation of Equation (2.6) is equivalent to a two-dimensional moving average filter for $I_{edge}(x, y, t)$ with filter coefficient $f(x, y; x_0, y_0, r, q)$. The pupil ellipse parameter center coordinate (x_0, y_0) and the radius r were estimated using the maximum value of the evaluation function $h(x_0, y_0, r, q; t)$. The parameters were written as $x_0(t)$, $y_0(t)$, and $r(t)$, respectively, and x_0 , y_0 , and r were functions of the frame t .

2.3.3 Three Steps Precision Improvement

In this research, approximating the pupil using an ellipse shape increased the number of parameters to be estimated and calculation cost, compared to the use of a circle shape. Therefore, we adopted a method for improving estimation accuracy consisting of three steps—rough, precise, and subpixel detection—to estimate the pupil center and radius mentioned in Section 2.3.2.

Initially, the rough detection estimation of the pupil center and radius from the entire image with an accuracy of 4 pixels was conducted. In order to detect a pupil with an accuracy of 4 pixels, the image $I(x, y, t)$ (after infrared spot filling) was spatially down-sampled by $1/4$. As a consequence, the search range r was also redefined as $r \in \{32/4, 36/4, \dots, 104/4\}$. Then, $x_0(t)$, $y_0(t)$, and $r(t)$ were estimated, using the method described in Section 2.3.2. Finally, these parameters were multiplied by four, in order to return them to the original scale.

Following this, the precise detection step used the estimated parameters $x_0(t)$, $y_0(t)$, and $r(t)$ from the rough detection step, in order to crop the search range. The cropped image was defined by the ranges $x_0(t) - r(t) - w \leq x \leq x_0(t) + r(t) + w$ and $y_0(t) - r(t) - w \leq y \leq y_0(t) + r(t) + w$, where w is the width of the area included around the

pupil. In this research, $w = 20$ pixels were selected as the included area width. In the rough pupil detection step, the pupil center (x_0, y_0) and radius r were estimated with an accuracy of 4 pixels. Therefore, in the precise pupil detection step, the search ranges for the pupil center (x_0, y_0) and radius r were limited to $x_0 \in \{x_0(t) - 4, x_0(t) - 3, \dots, x_0(t) + 3, x_0(t) + 4\}$, $y_0 \in \{y_0(t) - 4, y_0(t) - 3, \dots, y_0(t) + 3, y_0(t) + 4\}$, and $r \in \{r(t) - 4, r(t) - 3, \dots, r(t) + 3, r(t) + 4\}$. Other processes in this step were similar to those of the rough pupil detection step, in terms of estimating the pupil center (x_0, y_0) and radius r for each frame t . The method described in Section 2.3.2 was used to re-estimate the parameter with an accuracy of 1 pixel. The result of the estimation was defined by $x_0(t)$, $y_0(t)$, and $r(t)$.

Finally, in the subpixel detection step, the search range was further limited, using the parameters that were estimated in the precise detection step. The method described in Section 2.3.2 was used again, in order to re-estimate the parameters with an accuracy of 1/4 pixels. The search ranges for the pupil center (x_0, y_0) and radius r were limited to $x_0 \in \{x_0(t) - 1, x_0(t) - 0.75, \dots, x_0(t) + 0.75, x_0(t) + 1\}$, $y_0 \in \{y_0(t) - 1, y_0(t) - 0.75, \dots, y_0(t) + 0.75, y_0(t) + 1\}$, and $r \in \{r(t) - 1, r(t) - 0.75, \dots, r(t) + 0.75, r(t) + 1\}$.

2.3.4 Estimation of the Optimal Flatness Parameter q

According to the proposed method described in previous sections, the waveforms of the center coordinates $x_0(t)$, $y_0(t)$ and radius $r(t)$ of the pupil were estimated for each flatness parameter $q \in \{0.90, 0.95, 1.00, 1.05, 1.10\}$. The magnitude of the fluctuation of the radius $r(t)$ can be used as a measure of estimation accuracy—that is, the best selection for the flatness parameter—as the radius $r(t)$ does not change much, even if the center coordinates $x_0(t)$, $y_0(t)$ vary with nystagmus. Therefore, the optimum flatness parameter q is defined as the value that minimizes the magnitude of fluctuation of the radius $r(t)$. Figure 2.5 shows examples of the radius $r(t)$ estimated with each $q \in \{0.90, 0.95, 1.00, 1.05, 1.10\}$ for the same eye video. It can be concluded that $q = 0.95$ was optimal, as the radius $r(t)$ had minimum fluctuation. The specific calculation method for the magnitude of fluctuation is summarized in Appendix B.

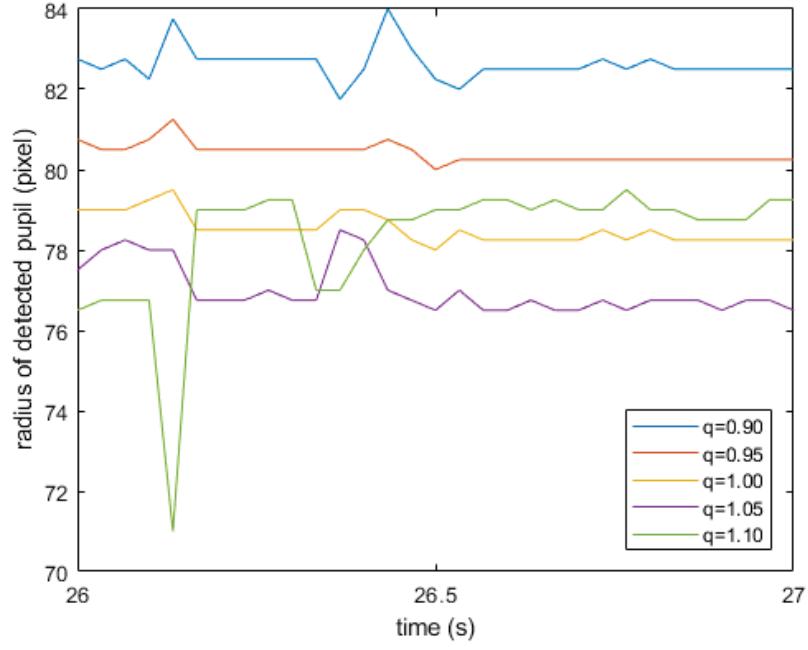


Figure 2.5: Sample of $r(t)$ for varying values of q .

2.4 Results

The existence of infrared spots influences the edge detection process for detecting the pupil contour, based on the intensity gradient between the pupil and the iris. Removing the spots is essential, as they decrease the accuracy of pupil detection. Due to the spots in the eye image, the edge detection step will also discern another circular border inside the pupil area. Consequently, when calculating the degree of similarity between the ellipse pattern and the edge image, the circular border from the spots shifts the pupil's estimated center. Figure 2.6 shows a comparison of edge detection results with and without the infrared spot filling process.

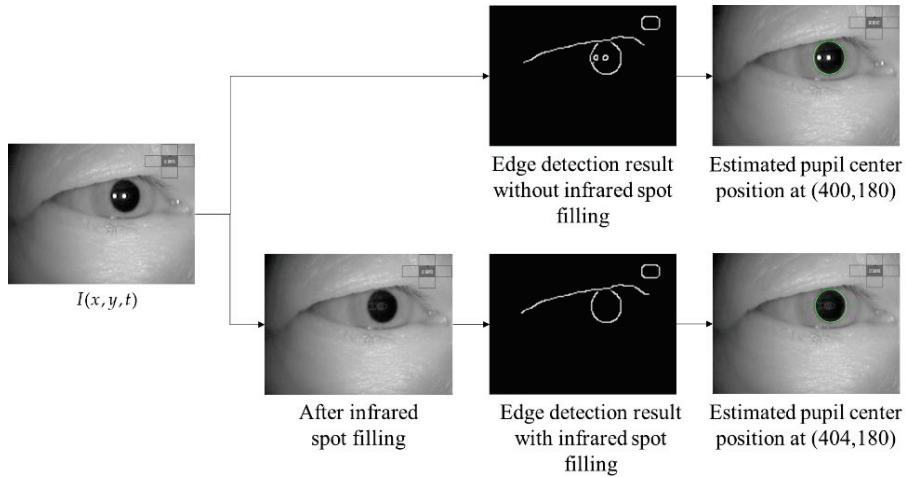


Figure 2.6: Sample of infrared spot filling to the detection result.

Figure 2.7 shows comparison results from the three-step precision improvement process described in Section 2.3.3. It can be observed that the nystagmus waveform becomes smoother at each step, due to the improvement of the pixel-order estimation. The pixel-order estimation is improved from 4 pixels to 1 pixel, and then to 1/4 pixel, as highlighted by the red ellipse. Figure 2.8 shows a sample of a nystagmus waveform generated by the proposed method. The waveform represents the pupil center position, based on its horizontal and vertical movement.

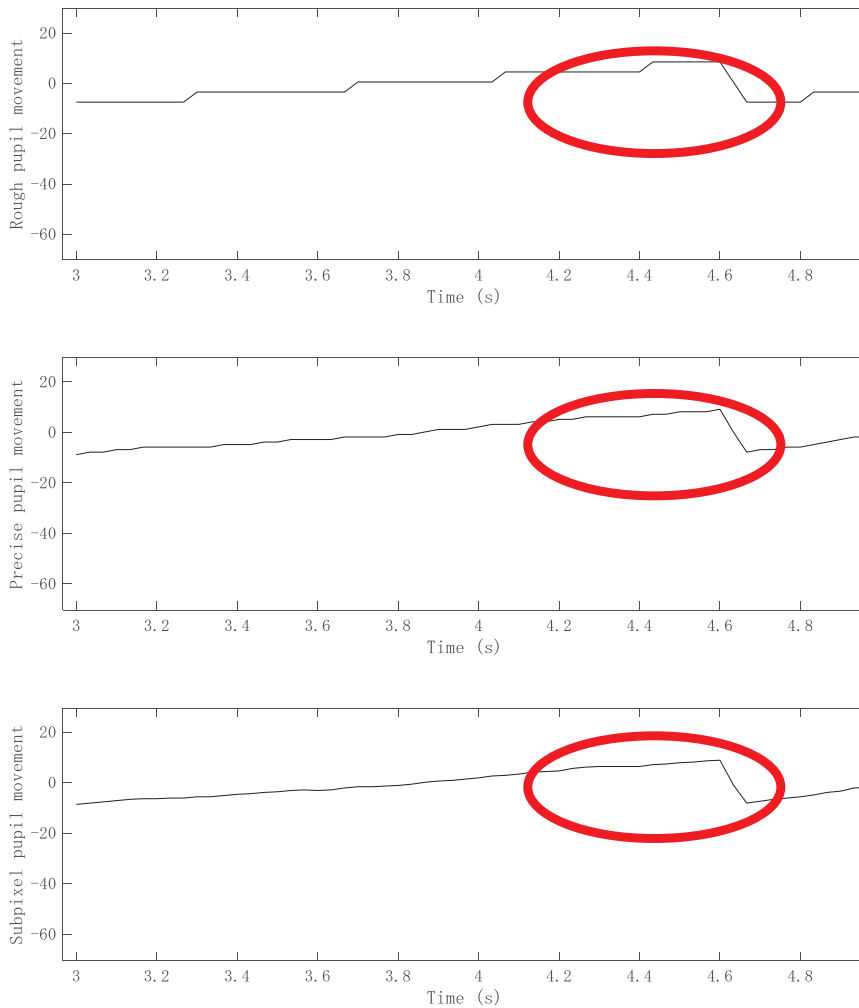


Figure 2.7: Comparison of rough, precise, and subpixel detection results.

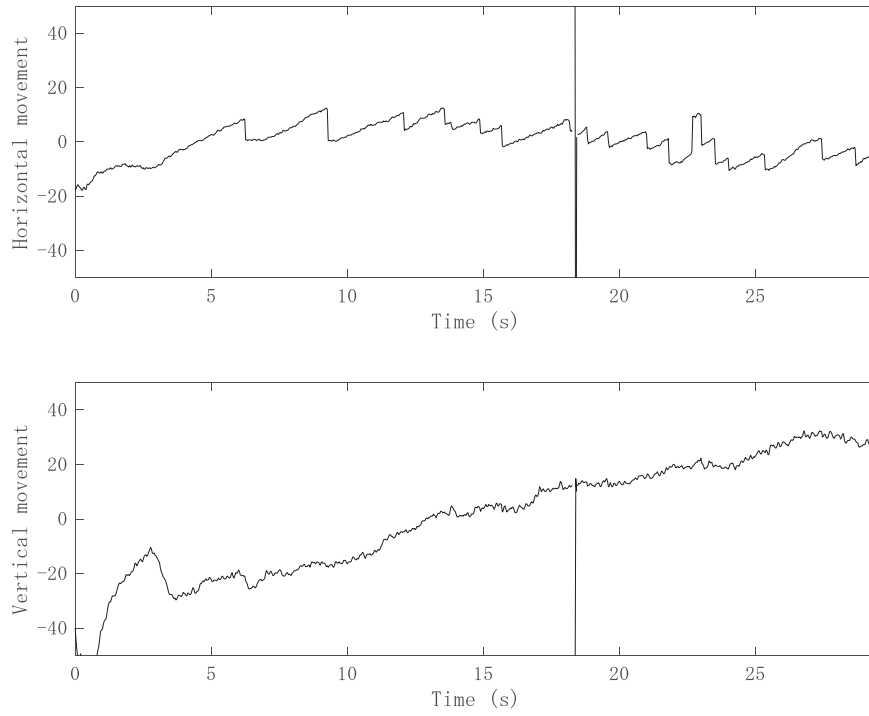


Figure 2.8: Sample of the nystagmus waveform generated by the proposed method.

2.5 Evaluation

2.5.1 Performance Evaluation for Partially Shown Pupil

As was highlighted in Section 1.1, patients who complain of dizziness often have difficulties in keeping their eyes open, which may require nystagmus to be measured from a semi-open state. Therefore, the performance of the proposed method was evaluated for eye movement videos under the condition that the video only shows a partial part of the pupil. Therefore, the video was cropped to show 100% to 10% of the pupil, with a gradual decrement by 10%. In this research, the removal of the pupil part started from the top area of the pupil. Figure 2.9 shows an illustration of pupil cropping.

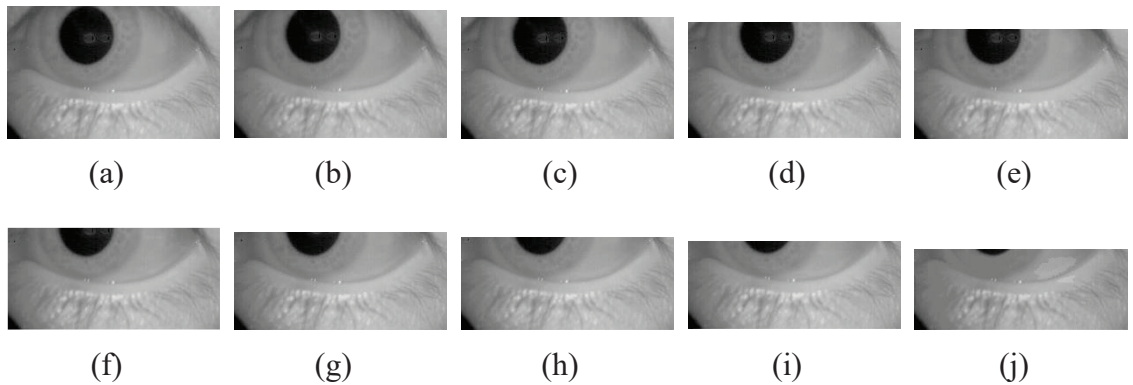


Figure 2.9: Illustration of pupil cropping: (a) 100%; (b) 90%; (c) 80%; (d) 70%; (e) 60%; (f) 50%; (g) 40%; (h) 30%; (i) 20%; and (j) 10%.

We calculated the Mean Square Error (MSE) between pupil position from a cropped pupil and fully visible pupil to assess the accuracy of the method. Based on visual observations, the obtained pupil center results for both methods had some outlier detections. In order to consider the outliers, outlier detection was not be included in the MSE calculation if the difference in pupil center position was equal to or larger than 20 pixels.

The MSE for all video frames was calculated as

$$MSE = \frac{1}{T} \sum_{t=1}^T \left(\left(\frac{x_0(t) - x_0'(t)}{N_x} \right)^2 + \left(\frac{y_0(t) - y_0'(t)}{N_y} \right)^2 \right), \quad (2.7)$$

where $(x_0(t), y_0(t))$ and $(x_0'(t), y_0'(t))$ are the pupil center positions in the videos with whole pupils and partial pupils, respectively.

We evaluated the performance of the proposed method in comparison to that of the conventional Hough transform method. For the evaluation, the MSE of each video is averaged, in order to obtain the mean MSE for each percentage of the visible pupil.

$$\overline{MSE} = \frac{1}{V} \sum_{v=1}^V MSE(v), \quad (2.8)$$

where $MSE(v)$ is the MSE from video number $v \in \{1, 2, \dots, V\}$, where V defines the total number of videos. Figure 2.10 shows the comparison results as a bar graph. In general, the Mexican hat-type ellipse pattern matching achieved a lower MSE , compared to the conventional Hough transform method. Specifically, if we define the acceptable range of error limit tolerance as 0.5 MSE , the performance of the proposed method achieved MSE values below the 0.5 limit until 20% of the pupil was visible. In other words, the proposed method can detect and track the movement of the center of the pupil almost as accurately as when 100% of the pupil is visible. In comparison, the conventional Hough transform method indicated a low MSE value under the 0.5 limit if only 90% of the pupil was visible. If the pupil was occluded more than 20%, the MSE value of the conventional Hough transform method increased significantly.

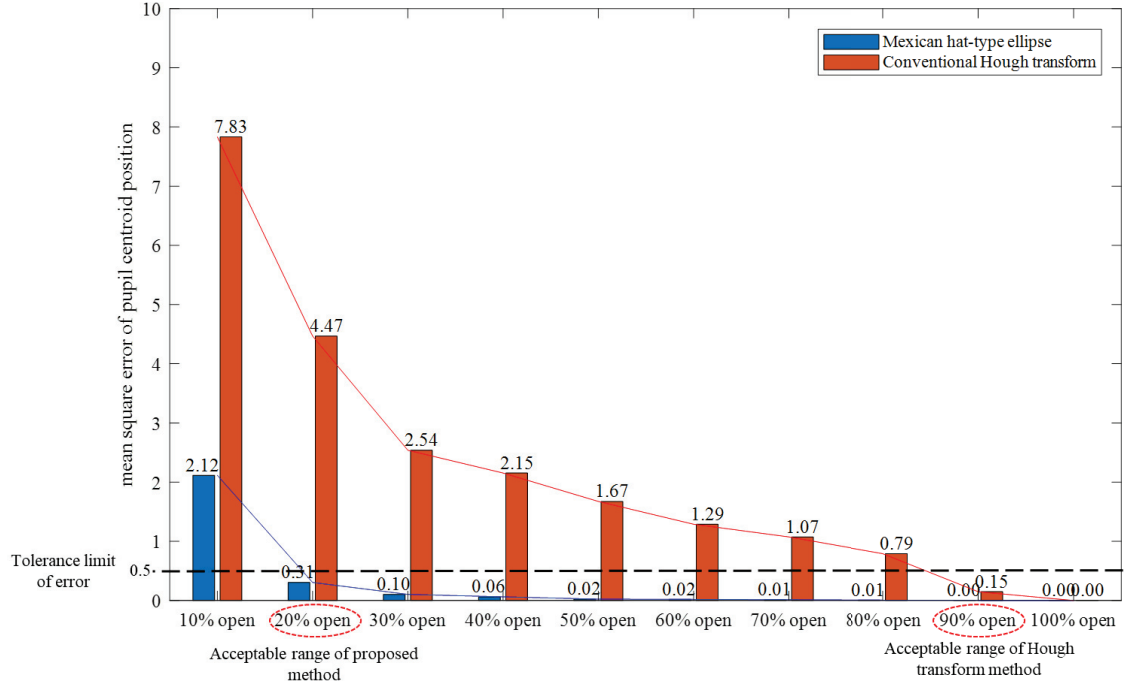
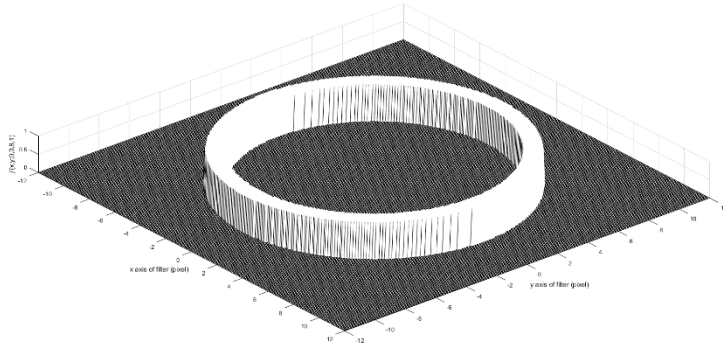
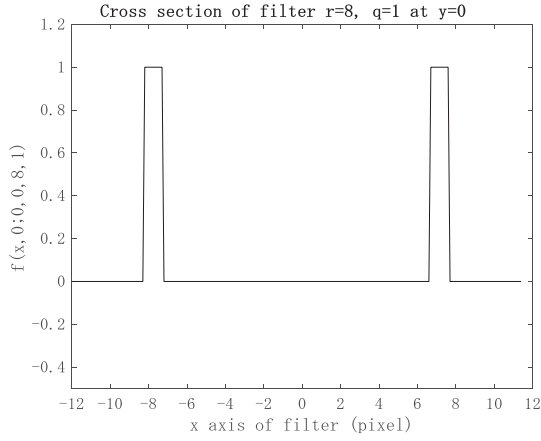


Figure 2.10: Comparison of MSE calculation results from the proposed Mexican hat-type ellipse pattern matching and the conventional Hough transform method.

The reason why the proposed method achieved higher estimation accuracy than the conventional Hough transform is described as follows. In the conventional Hough transform, the pixels within a certain width range are aggregated with equal weight for the target shape. Then, the maximum aggregate is used to estimate the parameters of the target shape. Therefore, circle detection by the conventional Hough transform is equivalent to pattern matching using a pattern with a uniform weight pattern, as shown in Figure 2.11. However, if the target shape has a blurry boundary that is not always clear, such as a whole pupil, the maximum degree of similarity $h(x_0, y_0, r, q; t)$ cannot be achieved, thus deteriorating the estimation accuracy. Therefore, we calculated the similarity degree $h(x_0, y_0, r, q; t)$ using the Mexican hat-type ellipse pattern, as shown in Figure 2.4a. The proposed method generates a sharp peak for boundary detection. Thus, it is expected to improve estimation accuracy.



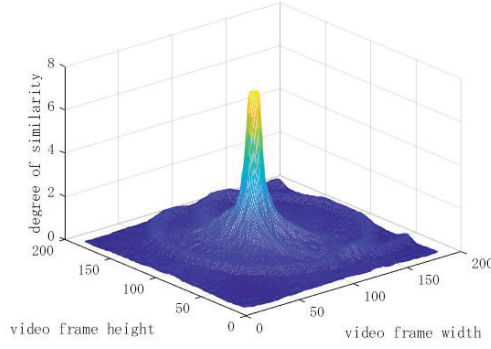
(a)



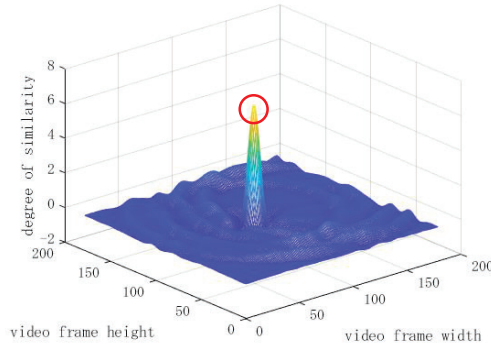
(b)

Figure 2.11: The example of the conventional Hough transform pattern with a uniform-valued ring: (a) Bird's-eye view; and (b) cross-section at $y = 0$.

Figure 2.12 shows the comparison result of the evaluation function $h(x_0, y_0, r, q; t)$ for the conventional Hough transform and the Mexican hat-type ellipse pattern. Figure 2.12a shows that the conventional Hough transform resulted in a flat peak, with some peaks resulting in the same degree of similarity. Therefore, it could not lead to a single maximum value of $h(x_0, y_0, r, q; t)$, representing the pupil center position. Meanwhile, the proposed Mexican hat-type pattern resulted in a single maximum peak value. Figure 2.12b shows the maximum peak, highlighted as a red circle, as the candidate for the pupil center.



(a)



(b)

Figure 2.12: The difference in peak sharpness for the evaluation function $h(x_0, y_0, r, q; t)$: (a) Conventional Hough transform; and (b) Mexican hat-type ellipse pattern.

As the performance of the proposed method is reliant on the detected pupil's shape, any artifacts that distort the pupil shape, such as accidents and optical diseases, will influence the results. For example, pupil abnormalities caused by Colobomas, Adie syndrome, or severe Uveitis can influence the accuracy of pupil tracking. Cloudiness in the cornea, such as Glaucoma and Cataracts, will also influence the accuracy of pupil tracking. Recommendations for further research include Nystagmus estimation for this abnormal and distorted pupil shape.

2.5.2 Performance Evaluation for Partially Shown Pupil

Using Equations (2.7) and (2.8), pupil center information from the proposed method was compared with the ground truth of the LPW data set. Using a similar approach, the performance of the conventional Hough transform method was also calculated. The proposed method achieved an MSE of 1.47, while the conventional Hough transform method achieved an MSE of 9.53.

2.5.3 Medical Specialist Validation

In this research, the Mexican hat-type ellipse pattern matching for detecting the pupil center was also evaluated using an expert validation approach. The expert validation approach was conducted by asking three medical specialists to evaluate the nystagmus waveform obtained from the proposed method. Then, the medical specialists wrote their reviews, regarding what the waveform represented. The medical specialist also commented on the eye movement video conditions and mentioned challenges in diagnosing the nystagmus state of disease.

Based on the medical specialists' reviews, the nystagmus waveform from the proposed method was evaluated clinically. The waveform could be used to assess unstable nystagmus without any problem. The proposed method can also detect the correct direction of the nystagmus case, and the detection was also accurate for both rapid and slow phases of nystagmus.

For example, the medical specialists highlighted the slow phase component of nystagmus in the horizontal direction of Video No. 1. This slow phase component is shown in Figure 2.13 as a nystagmus waveform generated by the proposed method. The medical specialist noticed that even the velocity of the slow phase was unstable; however, the system can be used to evaluate the nystagmus. In addition, as vertical nystagmus was not observed in the video, the slow phase was also undetected in the pupil vertical movement waveform, as shown in Figure 2.14.

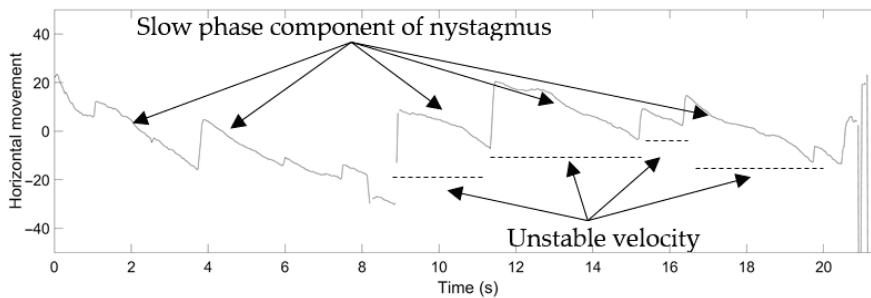


Figure 2.13: Nystagmus waveform from the proposed method for Video No. 1, horizontal movement of the pupil.

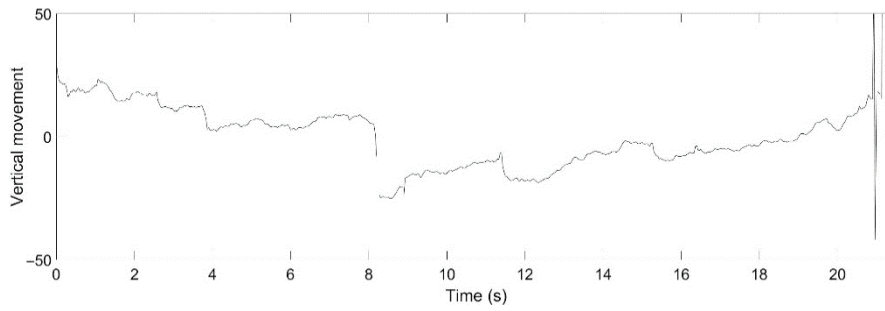


Figure 2.14: Nystagmus waveform from the proposed method for Video No. 1, vertical movement of the pupil.

In the case of nystagmus with high frequency, the proposed method could accurately capture the nystagmus. Furthermore, in the case of a low frequency of nystagmus, which is difficult to evaluate with the naked eye, it could be confirmed and detected in the waveform. An example of this can be seen in the nystagmus waveform for Video No. 28, as shown in Figure 2.15. The small amplitude of nystagmus was captured well by the proposed method for rapid and slow phase components in horizontal pupil movement.

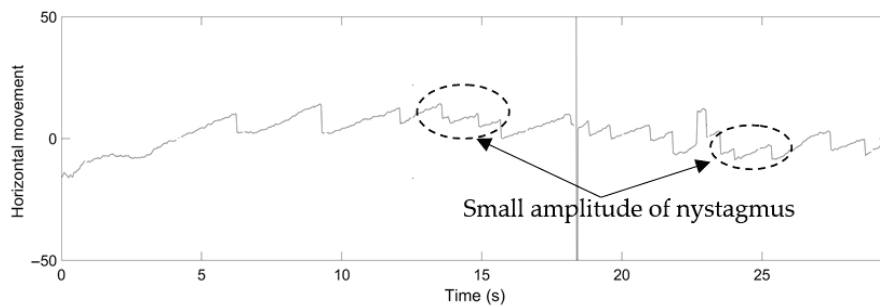


Figure 2.15: Nystagmus waveform from the proposed method for Video No. 28, horizontal movement of the pupil.

While the performance of the proposed method was well-recognized with a wide eyelid gap, the medical specialist also agreed that the waveform can be used to confirm nystagmus when the eyelid gap is narrow. The medical specialist mentioned that the condition of the narrow eyelid gap is difficult to evaluate. The entire iris is not visible in some videos, as some patients had difficulty in fully opening their eyes. However, the waveform can track pupil movement in both horizontal and vertical directions. The medical specialist mentioned that the waveform could still be used when up to 30% of the pupil was shown. For example, the medical specialist mentioned that the patient had difficulty opening her eyes in Video No. 2. Figure 2.16 shows a video frame from Video No. 2, which represents this condition. Figure 2.17a shows the nystagmus waveform that was obtained from Video No. 2. Based on this waveform, the vertical component of the

nystagmus was well-captured by the proposed method. In comparison, Figure 2.17b shows the nystagmus waveform from the conventional Hough transform method. The waveform had a high vibration of the vertical component of the nystagmus, due to the problem illustrated in Figure 2.12.

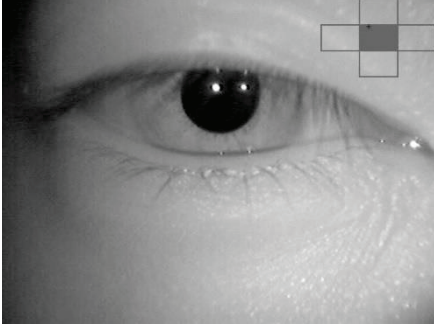


Figure 2.16: Sample of a video frame from Video No. 2.

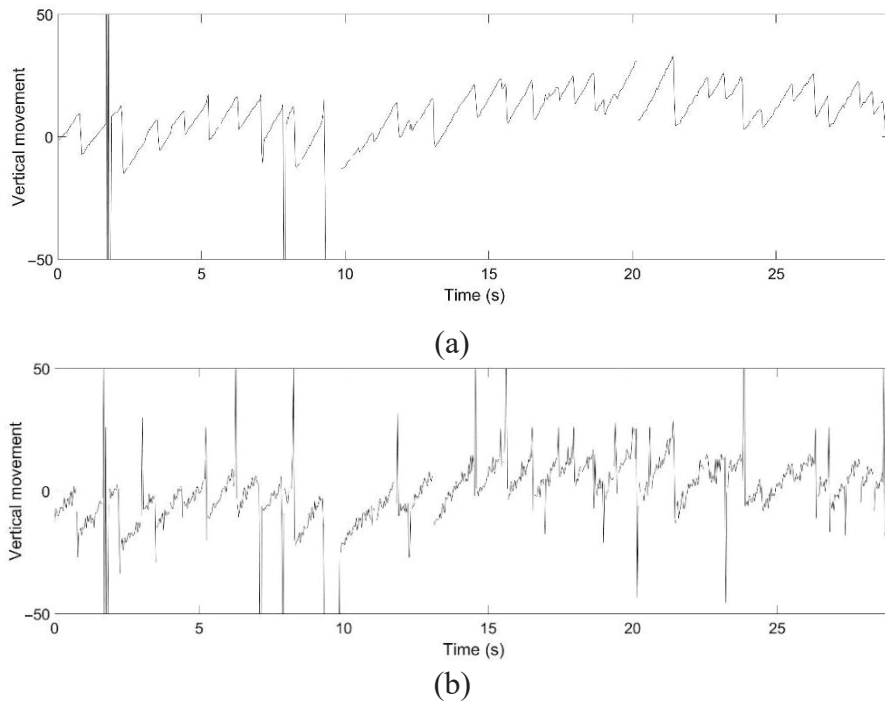


Figure 2.17: Nystagmus waveform for Video No. 2: (a) Using the proposed method and; (b) using the conventional Hough transform method.

In addition, the presence of contact lenses in the video does not affect the performance of the proposed method. Figure 2.18 shows a sample of a video frame from Video No. 11 which represents this condition, while Figure 2.19 shows a waveform that captures the horizontal rapid and slow phases of nystagmus for Video No. 11.

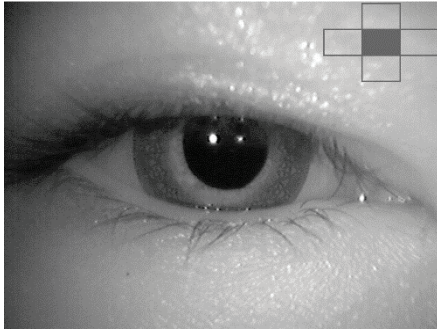


Figure 2.18: Sample of a video frame from Video No. 11.

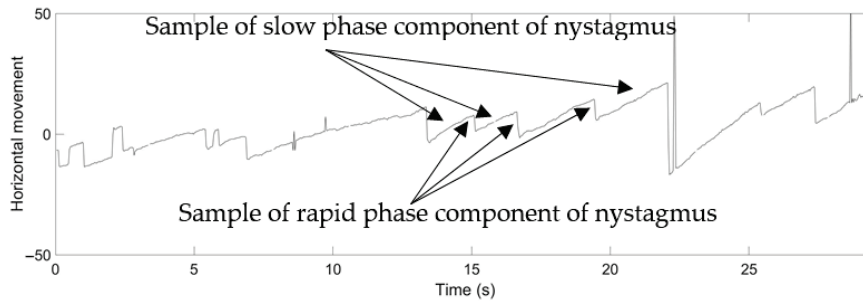


Figure 2.19: Nystagmus waveform from the proposed method for Video No. 11, horizontal movement of the pupil.

The medical specialist also recommended improving the infrared camera's specifications, as there was a limit, in terms of capture capacity, which prevented accurate evaluation of the rapid phase of nystagmus. The medical specialist also mentioned that the rotational component of nystagmus should be included in the waveform. Details of the medical specialists' review are provided in Appendix A, Table A2.

2.6 Conclusion

The principal purpose of this research was successfully achieved. Mexican hat-type ellipse pattern matching for detecting the center of a partially open pupil was proposed. Experiments using the implemented method on 37 eye videos were evaluated. The Mexican hat-type ellipse pattern matching approach achieved better performance, compared to the conventional Hough transform method. The evaluation also showed the robust performance of the proposed method, even when only 20% of the pupil was shown. Further evaluation of the performance of the proposed method using the LPW data set also showed that it can achieve a lower MSE, compared to the conventional Hough transform method. A review by medical specialists also provided evidence that the proposed method can support their diagnosis in the case of a low frequency of nystagmus, which is difficult to evaluate with the naked eye. In addition, the waveform generated by

the proposed method can reproduce eye movement in horizontal and vertical directions under the conditions of a narrow eyelid gap, which is difficult to evaluate. Therefore, the contributions of this research could lead to reasoning and diagnostic improvement of medical specialists, in the case of nystagmus estimation for dizziness diagnosis.

3 EARLY DETECTION AND TRACKING OF DISTANT INCOMING TRAFFIC USING IMPROVED DETECTION ON ROAD VANISHING POINT REFERENCE FOR ADAPTIVE TRAFFIC LIGHT SIGNALING

3.1 Materials and Methods

3.1.1 Video Test Material

Video is obtained from In-Luck Company that provides security business service such as traffic guidance in road construction site. In-luck Company positions camera on roadside to record daylight traffic activity. Recorded video from the camera shows perspective projection that makes far traffic from the camera looks small. Figure 3.1 shows sample of captured scene from twelve videos that are used as video test material.

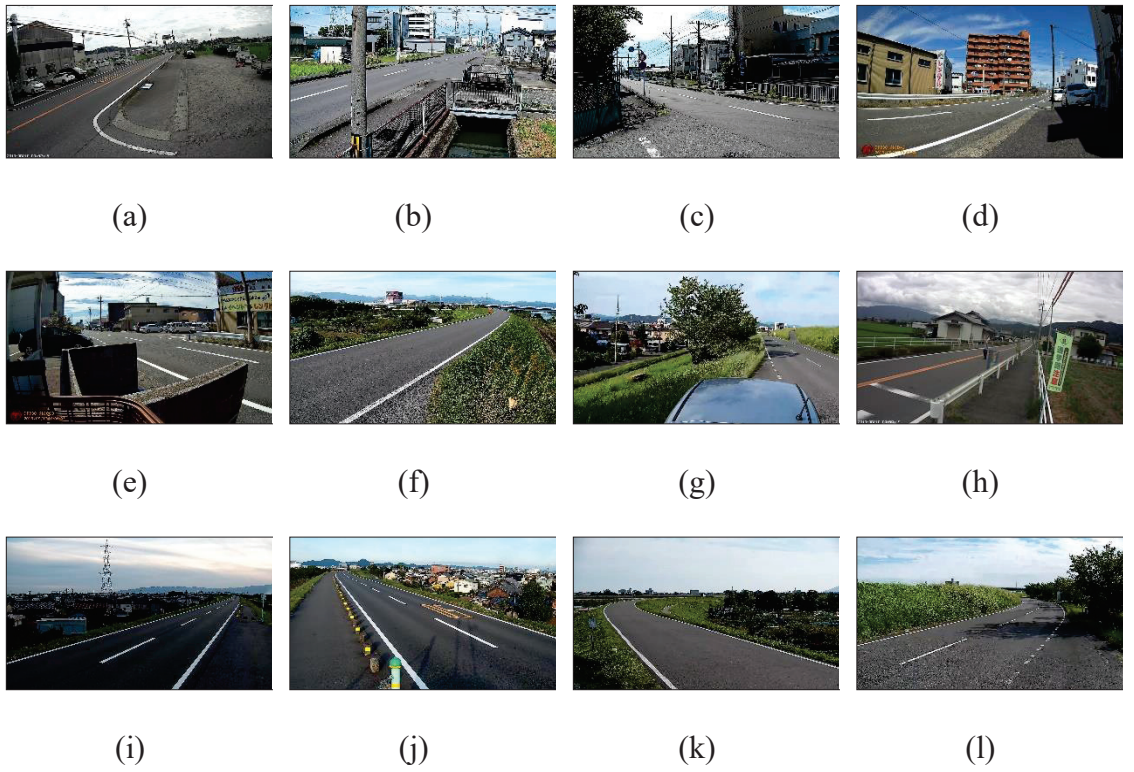


Figure 3.1: Sample of captured scene from video test material: (a) Site 1; (b) Site 2; (c) Site 3; (d) Site 4; (e) Site 5; (f) Site 6; (g) Site 7; (h) Site 8; (i) Site 9; (j) Site 10; (k) Site 11; (l) Site 12.

Each video frame is represented as $I(x, y, t) \in \{0, 1, \dots, 255\}$ where $x \in \{1, 2, \dots, N_x\}$, $y \in \{1, 2, \dots, N_y\}$, and $t \in \{1, 2, \dots, T\}$ with N_x and N_y are width and height of video frame. In

this study, the video test material has $N_x=1920$ and $N_y=1080$. T is total video frames that is calculated as

$$T = Vduration * Vfps, \quad (3.1)$$

where $Vduration(s)$ and $Vfps$ (frame/s) are duration of the video and video frame rate, respectively. $Vfps$ is 20 frame/s for Site 1 and Site 8 while $Vfps$ is 30 frame/s for the other videos.

$I(x, y, t)$ is a vector that comprises of three channels of RGB color expressed as

$$I(x, y, t) = \begin{cases} I_R(x, y, t) \\ I_G(x, y, t) \\ I_B(x, y, t) \end{cases} \quad (3.2)$$

where $I_R(x, y, t)$, $I_G(x, y, t)$, and $I_B(x, y, t)$ are channel of red, green, and blue color, respectively. In this study, $I(x, y, t)$ is converted to the grayscale image. Grayscale conversion is calculated as

$$I(x, y, t) = \frac{I_R(x, y, t) + I_G(x, y, t) + I_B(x, y, t)}{3}, \quad (3.3)$$

where $I(x, y, t)$ is grayscale conversion result that is used as input of the proposed method.

3.1.2 Proposed Method

Figure 3.2 shows the design of the proposed method and the following section describes details of the steps that are used in the proposed method.

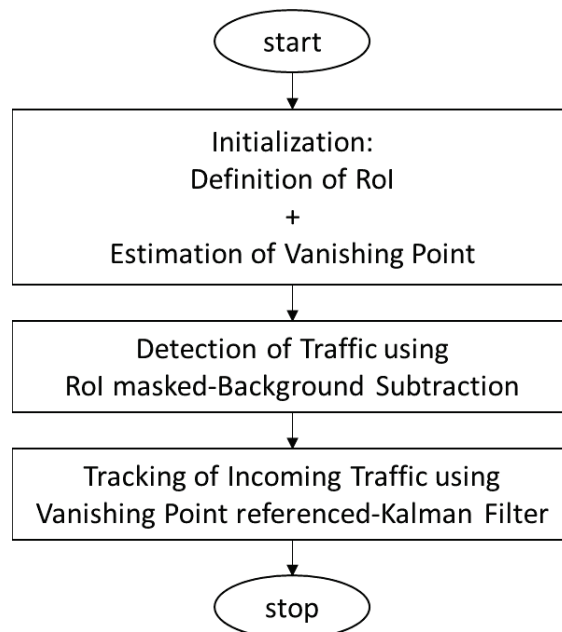


Figure 3.2: Design of proposed method.

3.1.2.1 Initialization Step

This study designs the initialization step as a preparation before conducting detection and tracking the incoming traffic. The first process in initialization step is defining the RoI. For RoI definition process, sequence of $I(x, y, t)$ with existence of traffic is used for RoI definition process. To obtain main movement of traffic with optimum processing time, this study recommends to down sampling $I(x, y, t)$ for $t \in \{Vfps, 2Vfps, \dots, T/Vfps\}$. This process is started by detecting foreground object using background subtraction method. Background frame for background subtraction process is calculated as

$$I_{BG}(x, y, t) = \text{median}(I(x, y, t - r), \dots, I(x, y, t), \dots, I(x, y, t + r)), \quad (3.4)$$

where $I_{BG}(x, y, t)$ is generated background and r is range of sequential frame that is calculated using median. Median calculation is used to minimize brightness fluctuation from sunlight intensity variation. In this step, r is predefined as 150 frames based on visual observation of detection quality. Then, background subtraction is calculated as

$$I_{FO}(x, y, t) = |I(x, y, t) - I_{BG}(x, y, t)|, \quad (3.5)$$

where $I_{FO}(x, y, t)$ is detected foreground from background subtraction calculation. The detected foreground still includes non-vehicle object and noise from variation of road texture and shadow. For this reason, noise removal process using thresholding is required. To remove the noise, $I_{FO}(x, y, t)$ with low-intensity value lower than th is removed from $I_{FO}(x, y, t)$ where th defined the threshold value. The selection of th is predefined in this study for each of video test material. Removal of this type of noise is calculated as

$$I_{FO}(x, y, t) = \begin{cases} 1, & I_{FO}(x, y, t) > th \\ 0, & \text{otherwise} \end{cases}, \quad (3.6)$$

where $I_{FO}(x, y, t)$ is redefined as result after noise filtering to avoid the complexity of notation. In addition, morphology process, including blurring and dilation, is also applied to remove the noise.

Since the purpose of RoI is to minimize unpredictable environment around roadway, this study localizes the main region where the vehicle movement exists. In order to obtain the RoI, this study uses frame difference method to obtain foreground object that has significant movement. Frame difference is calculated as

$$I_{MFO}(x, y, t) = |I_{FO}(x, y, t) - I_{FO}(x, y, t + 1)|, \quad (3.7)$$

where $I_{MFO}(x, y, t)$ is detected moving foreground object. Then, $I_{MFO}(x, y, t)$ from all t is summed up into one frame that shows number of region where moving foreground

object existed. In addition, thresholding is also applied to remove region with non-vehicle movement. The process is calculated as

$$I_{RoI}(x, y) = \begin{cases} 1, & th_{low} > \sum_{t=1}^T I_{MFD}(x, y, t) > th_{high}, \\ 0, & otherwise \end{cases} \quad (3.8)$$

where $I_{RoI}(x, y)$ represent the RoI. th_{low} and th_{high} are minimum and maximum threshold to filter non-vehicle movement. th_{low} and th_{high} are varied for each of video material and predefined based on visual observation.

The second process in initialization step is estimation of vanishing point coordinate. The vanishing point in this step is estimated from single $I(x, y, t)$ with t is preselected manually by visual observation. This study recommends to select t without existence of traffic to obtain precise vanishing point. In this process, $I_{VP}(x, y)$ is $I(x, y, t)$ for the preselected t . In addition, $I_{VP}(x, y)$ is also convolve with median filter with size of 5x5 pixels to reduce noise.

The estimation of vanishing point is started by calculating two components of WOD: differential excitation and orientation at each pixel location. Differential excitation is calculated based on difference between center pixel intensity and average intensity of all neighbors pixel in a $k \times k$ kernel size. Differential excitation is calculated using

$$\xi_{wod}(p_{center}) = \begin{cases} \sqrt{G(p_{center})}, & G(p_{center}) \geq 0, \\ 0, & otherwise \end{cases} \quad (3.9)$$

in which,

$$G(p_{center}) = \arctan\left(\frac{p_{center} - \overline{p_{neighbor}}}{p_{center}}\right), \quad (3.10)$$

where $\xi_{wod}(p_{center})$, p_{center} , $\overline{p_{neighbor}}$, and $G(p_{center})$ are differential excitation for p_{center} , intensity of center pixel, average intensity of all neighbors pixel, and the intensity difference, respectively. This study predefined $k=25$ as size of kernel.

$\xi_{wod}(p_{center})$ is further processed by thresholding to minimize the noise that exists in the frame texture features. This study defines $T=0.05$ as thresholding value. Normalized value of $\xi_{wod}(p_{center})$ that is larger than T is used to estimate the vanishing point.

This study uses Gabor filter to estimate dominant orientation at each of pixel location. Kernel of Gabor filter g that is centered at (x, y) for orientation φ_n and radial frequency $\omega = 2\pi/\lambda$ is defined as

$$g_{\varphi_n}(x, y) = e^{-\frac{1}{8\sigma^2}(4a^2+b^2)} \cdot (ia\omega - e^{c^2/2}), \quad (3.11)$$

where $a = x \cos \varphi_n + y \sin \varphi_n$ and $b = -x \sin \varphi_n + y \cos \varphi_n$. In this study, $\sigma = k/9$, $c=2.2$, and $\lambda = k\pi/10$ are a constant, similar to parameter setting in [33]. φ_n is calculated using

$$\varphi_n = \frac{(n-1)\pi}{N_\varphi} \quad n = 1, 2, \dots, N_\varphi, \quad (3.12)$$

where N_φ is total number of orientations. Dominant orientation for $I_{VP}(x, y)$ for each p_{center} is calculated using

$$\hat{I}_{\varphi_n}(p_{center}) = I_{VP}(p_{center}) * g_{\varphi_n}(p_{center}), \quad (3.13)$$

where $\hat{I}_{\varphi_n}(p_{center})$ is result of convolution between video frame and kernel of Gabor filter, and $*$ denotes convolution operator. $\hat{I}_{\varphi_n}(p_{center})$ as convolution result has real part and imaginary part. These two parts are used to calculate Gabor energy for each pixel in $\hat{I}_{\varphi_n}(p_{center})$. Gabor energy is calculated as

$$E_{\varphi_n}(p_{center}) = \sqrt{\text{Re}(\hat{I}_{\varphi_n}(p_{center}))^2 + \text{Im}(\hat{I}_{\varphi_n}(p_{center}))^2}, \quad (3.14)$$

where $E_{\varphi_n}(p_{center})$ is magnitude of Gabor energy at p_{center} . Finally, orientation at each of pixel location is defined as

$$\theta_{wod}(p_{center}) = \text{Argmax}_{\varphi_n} E_{\varphi_n}(p_{center}), \quad (3.15)$$

where $\theta_{wod}(p_{center})$ is p_{center} orientation.

The vanishing point is estimated based on result of Line-Voting Scheme (LVS). Firstly, LVS sets accumulator space with the same size as $I_{VP}(x, y)$ with initial zero value. Secondly, $\xi_{wod}(p_{center})$ and its counterpart $\theta_{wod}(p_{center})$ act as a voter that draws rays in the accumulator space. The corresponding accumulator space is increased by 1 if the rays lies over it. Finally, maximum value in the accumulator space is defined as vanishing point coordinate, (x_{vp}, y_{vp}) .

3.1.2.2 Object Detection using Background Subtraction

The second step uses defined RoI from the initialization step to mask the $I(x, y, t)$. The masking process conducted by

$$I_{masked}(x, y, t) = \begin{cases} I(x, y, t), & I_{RoI}(x, y) = 1 \\ 0, & \text{otherwise} \end{cases}, \quad (3.16)$$

where $I_{masked}(x, y, t)$ is the masking result. Then, similar process of background subtraction as used in Eq. (3.4-3.6) is applied for all frame in $I_{masked}(x, y, t)$. $I_{FO}(x, y, t)$ is obtained from the process.

3.1.2.3 Object Tracking using Kalman Filter

$I_{FO}(x, y, t)$ result from second step is further processed to track its movement. The process is to associate detected $I_{FO}(x, y, t)$ based on its movement from frame to frame. Since this study uses video from a stationary video camera, the Kalman filter [37] can predict object tracks in each frame and determine the likelihood of each detection to each track. Following this, track maintenance is also applied to update any new object or vanishing object from the video frame.

Mainly, the motion estimation process in this step follows Matlab documentation [38]. Configuration is applied for minimum detection of the blob area for 100 pixels to cope with the condition in detecting small resolution of incoming traffic in the study case.

Motion estimation step results coordinate of the detected object from each frame. However, because there is also a possibility to capture the outgoing traffic approaching the vanishing point, the result is filtered for detection that gets further from the vanishing point only.

Distance of vanishing point to detected object is calculated using Euclidean distance. Euclidean distance is calculated as

$$d(t) = \sqrt{(x_{obj}(t) - x_{vp})^2 + (y_{obj}(t) - y_{vp})^2}, \quad (3.17)$$

where $d(t)$ and $(x_{obj}(t), y_{obj}(t))$ are Euclidean distance of object and coordinate of detected object at frame t , respectively. Filtering incoming traffic is conducted by

$$incoming(t) = \begin{cases} 1, & d(t) > d(t-1) \\ 0, & otherwise \end{cases}, \quad (3.18)$$

where $incoming(t)$ is label for incoming traffic.

3.2 Result and Discussion

3.2.1 Initialization Step

Figure 3.3 shows result of the initialization step. The columns represent result of grayscale image $I_{VP}(x, y)$, grayscale image that has been masked with defined RoI, and overlaid vanishing point, respectively. The rows represent each of video test material.





Figure 3.3: Result of initialization step of the proposed method: (a) Site 1; (b) Site 2; (c) Site 3; (d) Site 4; (e) Site 5; (f) Site 6; (g) Site 7; (h) Site 8; (i) Site 9; (j) Site 10; (k) Site 11; (l) Site 12.

As shown in the second column of Figure 3.3, the defined RoI can cover main road lane where most of traffic activity is existing. The defined RoI is also affected by perspective projection that which one side of the RoI become smaller as the road lane getting distant. Even though some of RoI shape irregular, the objective to exclude movement of non-vehicle object that is unpredictable in the captured scene such as shrub, grass and tree, and flag can be minimized. This condition can be observed from second column of Figure 3.3: (f-l) where the road is bordered by grass.

The third column of Figure 3.3 shows estimated vanishing point from the initialization step as red crosshair. As a comparison, ground truth of the vanishing point is shown as green crosshair. In this study, the ground truth of vanishing point is manually marked with visual fixation.

Qualitative comparison shows that the estimated vanishing point from the initialization step almost can match the ground truth in straight road such as in Figure 3.3: (a-e). It is because the LVS mainly defines the vanishing point based on major orientation of straight edge object that exists in the $I_{VP}(x, y)$. Existence of road line, road fence, pavement, and aerial utility cable influences accuracy of the estimated vanishing point. In case of Figure 3.3: (f-k), curved shape of the roadway makes the estimated vanishing point shifted from the ground truth. In this condition, the curved edge object cannot get the major vote in the LVS. Observation of experiment result also shows that existence of straight edge of bridge and other visible roadway also makes the vanishing point is shifted in curved road.

In addition, Figure 3.3: (l) shows how the vanishing point is estimated in S-curved road. In this case, edge of curved road is not clearly visible. Figure 3.4 highlights how these objects (highlighted in as red ovals) influence the voting process of rays in accumulator space.

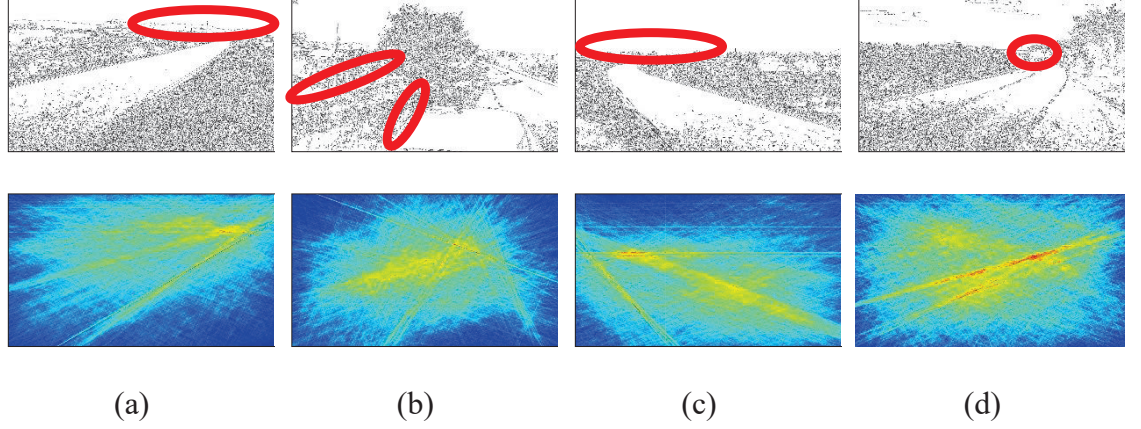


Figure 3.4: Sample detected edge (first row) and accumulator space (second row) that influence the estimated vanishing point: (a) Site 6; (b) Site 7; (c) Site 11; (d) Site 12.

As quantitative comparison, estimation error of the vanishing point is calculated using normalized Euclidean distance [39]. The estimation error is defined as

$$\delta = \frac{\sqrt{(x_{vp} - x_{gt})^2 + (y_{vp} - y_{gt})^2}}{\sqrt{N_x^2 + N_y^2}}, \quad (3.19)$$

where δ and (x_{vp}, y_{vp}) are estimation error and ground truth coordinate, respectively. δ near to 0 represents close estimation of the vanishing point to the ground truth; otherwise, δ near to 1 represents inaccuracy of estimation.

Table 3.1 tabulates δ for each video test material with variation of $N_\phi \in \{9, 12, 18, 36, 180\}$. Variation of N_ϕ influences how accurate the estimation of vanishing point based on the orientation in Gabor filter calculation. In general, error of estimation is relatively low with maximum value of 0.1635. Increasing N_ϕ means increasing resolution of ϕ_n precision. Thus, the error estimation can be lowered. However, the processing time for high resolution orientation is also increased.

Table 3.1 Quantitative comparison of vanishing point estimation error (δ) for variation of N_φ .

Site	δ					$\bar{\delta}$
	9	12	18	36	180	
1	0.0134	0.0204	0.0051	0.0040	0.0080	0.0111
2	0.0947	0.0501	0.0053	0.0027	0.0171	0.0369
3	0.1026	0.0161	0.0025	0.0040	0.0149	0.0288
4	0.0282	0.0188	0.0166	0.0087	0.0059	0.0151
5	0.0538	0.0074	0.0144	0.0099	0.0149	0.0208
6	0.0328	0.0301	0.0428	0.0174	0.0251	0.0312
7	0.0364	0.0163	0.0365	0.0352	0.0115	0.0242
8	0.0221	0.0148	0.0278	0.0154	0.0180	0.0195
9	0.0081	0.0199	0.0222	0.0246	0.0290	0.0205
10	0.1635	0.0278	0.0270	0.0283	0.0398	0.0587
11	0.0390	0.0946	0.0271	0.0375	0.0444	0.0500
12	0.0241	0.0067	0.0724	0.0132	0.0150	0.0259

3.2.2 Detection and Tracking of Incoming Traffic

Figure 3.5 shows sample result of the traffic detection from the second step. The first row shows detection of $I_{FO}(x, y, t)$ inside the RoI. To achieve the study purpose, small object that appear inside the RoI is categorized as candidate of the incoming traffic. Then, the detection is further processed by Kalman filter to calculate motion estimation. The second row of Figure 3.5 shows detected object that getting distant from the vanishing point. These objects are defined as final result of the proposed method.

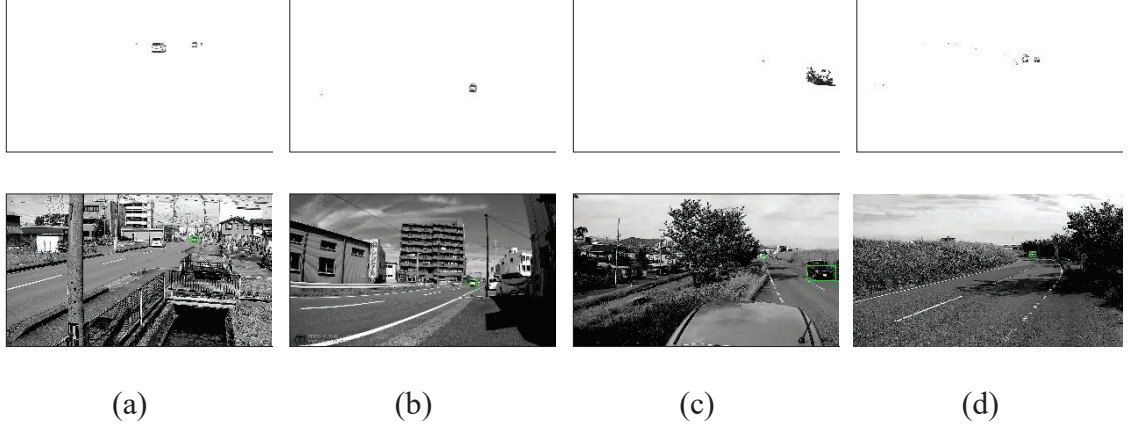


Figure 3.5: Sample detection of $I_{FO}(x, y, t)$ (first row) and final result of incoming traffic detection (second row): (a) Site 2; (b) Site 4; (c) Site 7; (d) Site 12

3.3 Evaluation

The performance of the proposed method is benchmarked to the performance of Regions with R-CNN [40] object detector. The evaluation is conducted based on how early the method can detect the incoming traffic. First, the CNN in R-CNN is created using image input layer, 2D convolution layer for Convolutional Neural Networks, Rectified linear unit (ReLU) layer, Max pooling layer, Fully connected layer, Softmax layer, and Classification output layer for a neural network.

R-CNN processed a CIFAR-10 data set that contains 50,000 training images that will be used to train a CNN. The training images have ten categories, including automobile, that suitable for the study case. The training is conducted using Stochastic Gradient Descent with Momentum (SGDM) with an initial learning rate of 0.001. The initial learning rate is reduced every eight epochs for a total of 40 epochs training.

After ensuring R-CNN is working well for the CIFAR-10 data set, the network is also trained using self-generated data set. The data set is generated based on the video frame from the first and the second camera. Then, each image is labeled manually based on visual observation from the frame. Forty vehicle images are selected from the video frames. The image shows the front part of the vehicle that represents incoming traffic. The image is also varied in terms of size, position in the roadway, and grayscale intensity. The training is conducted using the same Stochastic Gradient Descent with Momentum (SGDM) with an initial learning rate of 0.001 for 100 epochs training. The entire implementation of R-CNN used in this study follows Matlab documentation [41] with some modifications to accommodate study case conditions.

The benchmarking process is started by selecting sequential frames as object of evaluation. The sequential frames are selected manually from each the video test material, started at t_{start} . The sequential frames show the movement of incoming traffic from the ground truth vanishing point for duration of 10 seconds that is defined by t_{end} . The frontmost traffic is defined as detection target vehicle. These sequential frames are processed by the proposed method and R-CNN. The benchmarking process compares how early both methods detect the incoming traffic by confirming the detection result visually that is defined by t_{detect} .

Table 3.2 summarizes detection result for the benchmarking process. In general, the proposed method has earlier detection of incoming traffic than the R-CNN method. For twelve the video test materials in this study case, the proposed method requires average of 17.75 frames to detect the target vehicle while the R-CNN requires average of 63.36 frames to detect the target vehicle. For Site 9, R-CNN method cannot detect the target vehicle because until t_{end} , the target vehicle is too small to be detected. The result shows that R-CNN requires larger size vehicle image to ensure its recognition as a vehicle. For example, in Site 1, the vehicle is detected if its minimum size is 96x200 pixels. In Site 8, vehicle is detected if its minimum size is 141x176pixels. For Site 12, R-CNN method has earlier detection than the proposed method. It is because the detected vanishing point from the proposed method exists in the center of roadway due to S-curved characteristic of the roadway. As shown in Figure 3: (l), incoming car needs to pass the vanishing point first before can be detected as incoming traffic.

Table 3.2 Benchmarking result on detection of incoming traffic.

Site	t_{start}	t_{end}	t_{detect}	
			Proposed method	R-CNN
1	870	1070	880	890
2	360	660	370	437
3	120	420	139	202
4	4200	4500	4211	4276
5	1110	1410	1121	1321
6	600	900	644	672
7	1800	2100	1810	1817
8	1080	1280	1090	1120
9	3780	4080	3792	*
10	1080	1380	1094	1132
11	4500	4800	4538	4545
12	690	990	714	695

*until t_{end} , R-CNN cannot detect the target vehicle

3.4 Conclusions

Principal purpose of the research has been successfully achieved. Improved detection and tracking of distant incoming traffic based on vanishing point reference is proposed. It has been shown that RoI definition can be done and the vanishing point can be estimated in the proposed method. The proposed method would result earlier detection compared to the used of R-CNN. The findings also suggest that performance of the proposed method is dependent on the RoI definition and number of pixel orientation in Gabor filter calculation that influence accuracy of the vanishing point.

LIST OF PUBLICATIONS

1. Y. A. Syahbana, Y. Yasunari, M. Hiroyuki, A. Mitsuhiro, S. Kanade, and M. Yoshitaka, “Nystagmus Estimation for Dizziness Diagnosis by Pupil Detection and Tracking Using Mexican-Hat-Type Ellipse Pattern Matching,” *Healthcare*, vol. 9, no. 7, p. 885, Jul. 2021.
2. Y. A. Syahbana and Y. Yasunari, “Early Detection of Incoming Traffic for Automatic Traffic Light Signaling during Roadblock using Vanishing Point-Guided Object Detection and Tracking,” *The Society of Instrument and Control Engineers (SICE) Annual Conference 2021*, pp. 1463–1468, 2021.
3. Y. A. Syahbana and Y. Yasunari, “Detection of Congested Traffic Flow during Road Construction using Improved Background Subtraction with Two Levels RoI Definition,” *International Applied Business and Engineering Conference 2021*, pp. 71–76, 2021

REFERENCES

- [1] J. Baskaran and R. Subban, “Compressive Object Tracking – A Review and Analysis,” 2014.
- [2] S. M. Schappert and C. W. Burt, “Ambulatory care visits to physician offices, hospital outpatient departments, and emergency departments,” National Center for Health Statistics, 2006.
- [3] “Vertigo and Balance Disorders Q&A.”
<http://www.memai.jp/QandA/QandAenglish.htm>.
- [4] A. T. H. Lee Dr., “Diagnosing the cause of vertigo: A practical approach,” *Hong Kong Med. J.*, vol. 18, no. 4, pp. 327–332, 2012.
- [5] R. E. Post and L. M. Dickerson, “Dizziness: A diagnostic approach,” *Am. Fam. Physician*, vol. 82, no. 4, pp. 361–368, 2010.
- [6] D. E. Newman-Toker, L. M. Cannon, M. E. Stofferahn, R. E. Rothman, Y. H. Hsieh, and D. S. Zee, “Imprecision in patient reports of dizziness symptom quality: A cross-sectional study conducted in an acute care setting,” *Mayo Clin. Proc.*, vol. 82, no. 11, pp. 1329–1340, 2007, doi: 10.4065/82.11.1329.
- [7] H. R. Choi, S. Choi, J. E. Shin, and C. H. Kim, “Nystagmus findings and hearing recovery in idiopathic sudden sensorineural hearing loss without dizziness,” *Otol. Neurotol.*, vol. 39, no. 10, pp. e1084–e1090, 2018, doi: 10.1097/MAO.0000000000002005.
- [8] T. D. Fife *et al.*, “Practice Parameter: Therapies for benign paroxysmal positional vertigo (an evidence-based review): [RETIRED],” *Neurology*, vol. 70, no. 22, pp. 2067 LP – 2074, May 2008, doi: 10.1212/01.wnl.0000313378.77444.ac.
- [9] N. Bhattacharyya *et al.*, “Clinical Practice Guideline: Benign Paroxysmal Positional Vertigo (Update),” *Otolaryngol. - Head Neck Surg. (United States)*, vol. 156, no. 3_suppl, pp. S1–S47, 2017, doi: 10.1177/0194599816689667.
- [10] J. A. Edlow, “Diagnosing Patients With Acute-Onset Persistent Dizziness,” *Ann. Emerg. Med.*, vol. 71, no. 5, pp. 625–631, 2018, doi: 10.1016/j.annemergmed.2017.10.012.
- [11] A. A. Tarnutzer, A. L. Berkowitz, K. A. Robinson, Y. H. Hsieh, and D. E. Newman-Toker, “Does my dizzy patient have a stroke? A systematic review of

- bedside diagnosis in acute vestibular syndrome,” *Cmaj*, vol. 183, no. 9, pp. 571–592, 2011, doi: 10.1503/cmaj.100174.
- [12] A. A. Tarnutzer and D. Straumann, “Nystagmus,” *Curr. Opin. Neurol.*, vol. 31, no. 1, pp. 74–80, 2018, doi: 10.1097/WCO.0000000000000517.
- [13] R. J. Leigh and J. C. Rucker, “Nystagmus and Related Ocular Motility Disorders,” *Walsh Hoyt’s Clin. Neuro-Ophthalmology*, no. 2, pp. 1–89, 2005.
- [14] N. K. Macdonald, D. Kaski, Y. Saman, A. A. S. Sulaiman, A. Anwer, and D. E. Bamiau, “Central positional nystagmus: A systematic literature review,” *Front. Neurol.*, vol. 8, no. APR, pp. 1–11, 2017, doi: 10.3389/fneur.2017.00141.
- [15] R. J. Leigh and S. Khanna, “Neuroscience of Eye Movements,” *Adv. Clin. Neurosci. Rehabil.*, vol. 5, no. 6, pp. 12–15, 2006.
- [16] M. Strupp, O. Kremmyda, and T. Brandt, “Pharmacotherapy of vestibular disorders and nystagmus,” *Semin. Neurol.*, vol. 33, no. 3, pp. 286–296, 2013, doi: 10.1055/s-0033-1354594.
- [17] D. Ehrhardt and E. Eggenberger, “Medical treatment of acquired nystagmus,” *Curr. Opin. Ophthalmol.*, vol. 23, no. 6, pp. 510–516, 2012, doi: 10.1097/ICU.0b013e328358ba6e.
- [18] J. E. Self *et al.*, “Management of nystagmus in children: a review of the literature and current practice in UK specialist services,” *Eye*, vol. 34, no. 9, pp. 1515–1534, 2020, doi: 10.1038/s41433-019-0741-3.
- [19] R. F. Pilling, J. R. Thompson, and I. Gottlob, “Social and visual function in nystagmus,” *Br. J. Ophthalmol.*, vol. 89, no. 10, pp. 1278–1281, 2005, doi: 10.1136/bjo.2005.070045.
- [20] K. A. Kerber *et al.*, “Nystagmus assessments documented by emergency physicians in acute dizziness presentations: A target for decision support?,” *Acad. Emerg. Med.*, vol. 18, no. 6, pp. 619–626, 2011, doi: 10.1111/j.1553-2712.2011.01093.x.
- [21] M. Porta and A. Ravarelli, “Eye-based user interfaces: Some recent projects,” *3rd Int. Conf. Hum. Syst. Interact. HSI’2010 - Conf. Proc.*, pp. 289–294, 2010, doi: 10.1109/HSI.2010.5514555.
- [22] N. H. Cuong and H. T. Hoang, “Eye-gaze detection with a single webCAM based

- on geometry features extraction,” *11th Int. Conf. Control. Autom. Robot. Vision, ICARCV 2010*, no. February, pp. 2507–2512, 2010, doi: 10.1109/ICARCV.2010.5707319.
- [23] L. Kunhui, H. Jiyong, C. Jiawei, and Z. Changle, “Real-time eye detection in video streams,” *Proc. - 4th Int. Conf. Nat. Comput. ICNC 2008*, vol. 6, no. 2006, pp. 193–197, 2008, doi: 10.1109/ICNC.2008.278.
 - [24] T. Morav, “An Approach to Iris and Pupil Detection in Eye Images,” *XII Int. Ph.D. Work. OWD, Oct. 2010*, no. October, pp. 23–26, 2010.
 - [25] Z. H. Zhou and X. Geng, “Projection functions for eye detection,” *Pattern Recognit.*, vol. 37, no. 5, pp. 1049–1056, 2004, doi: 10.1016/j.patcog.2003.09.006.
 - [26] Dongheng Li, D. Winfield, and D. J. Parkhurst, “Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches,” in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05) - Workshops*, 2005, vol. 3, no. 34, pp. 79–79, doi: 10.1109/CVPR.2005.531.
 - [27] A. Al-Rahayfeh and M. Faezipour, “Eye tracking and head movement detection: A state-of-art survey,” *IEEE J. Transl. Eng. Heal. Med.*, vol. 1, no. March 2014, pp. 11–22, 2013, doi: 10.1109/JTEHM.2013.2289879.
 - [28] S. Araghi, A. Khosravi, and D. Creighton, “A review on computational intelligence methods for controlling traffic signal timing,” *Expert Syst. Appl.*, vol. 42, no. 3, pp. 1538–1550, 2015, doi: 10.1016/j.eswa.2014.09.003.
 - [29] F. Lian, B. Chen, K. Zhang, L. Miao, J. Wu, and S. Luan, “Adaptive traffic signal control algorithms based on probe vehicle data,” *J. Intell. Transp. Syst. Technol. Planning, Oper.*, vol. 25, no. 1, pp. 41–57, 2021, doi: 10.1080/15472450.2020.1750384.
 - [30] M. Koziarski and B. Cyganek, “Impact of low resolution on image recognition with deep neural networks: An experimental study,” *Int. J. Appl. Math. Comput. Sci.*, vol. 28, no. 4, pp. 735–744, 2018, doi: 10.2478/amcs-2018-0056.
 - [31] L. Xue, X. Zhong, R. Wang, J. Yang, and M. Hu, “Low - resolution vehicle recognition based on deep feature fusion,” *Multimed. Tools Appl.*, vol. 77, no. 20, pp. 27617–27639, 2018, doi: 10.1007/s11042-018-5940-6.
 - [32] W. Zou and P. C. Yuen, “Very low resolution face recognition problem,” *IEEE 4th*

Int. Conf. Biometrics Theory, Appl. Syst. BTAS 2010, no. May, 2010, doi: 10.1109/BTAS.2010.5634490.

- [33] W. Yang, X. Luo, B. Fang, D. Zhang, and Y. Y. Tang, “Fast and accurate vanishing point detection in complex scenes,” *2014 17th IEEE Int. Conf. Intell. Transp. Syst. ITSC 2014*, pp. 93–98, 2014, doi: 10.1109/ITSC.2014.6957672.
- [34] M. Tonsen, X. Zhang, Y. Sugano, and A. Bulling, “Labelled pupils in the wild: A dataset for studying pupil detection in unconstrained environments,” *Eye Track. Res. Appl. Symp.*, vol. 14, pp. 139–142, 2016, doi: 10.1145/2857491.2857520.
- [35] S. Y. Han, H. J. Kwon, Y. Kim, and N. I. Cho, “Noise-Robust Pupil Center Detection through CNN-Based Segmentation with Shape-Prior Loss,” *IEEE Access*, vol. 8, pp. 64739–64749, 2020, doi: 10.1109/ACCESS.2020.2985095.
- [36] N. H. Lestriandoko and R. Sadikin, “Circle detection based on hough transform and Mexican Hat filter,” *Proceeding - 2016 Int. Conf. Comput. Control. Informatics its Appl. Recent Prog. Comput. Control. Informatics Data Sci. IC3INA 2016*, no. 1, pp. 153–157, 2017, doi: 10.1109/IC3INA.2016.7863041.
- [37] G. Welch and G. Bishop, “An Introduction to the Kalman Filter,” *In Pract.*, vol. 7, no. 1, pp. 1–16, 2006, doi: 10.1.1.117.6808.
- [38] Matlab, “Motion-Based Multiple Object Tracking.” <https://de.mathworks.com/help/vision/ug/motion-based-multiple-object-tracking.html>.
- [39] P. Moghadam, J. A. Starzyk, and W. S. Wijesoma, “Fast vanishing-point detection in unstructured environments,” *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 425–430, 2012, doi: 10.1109/TIP.2011.2162422.
- [40] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 580–587, 2014, doi: 10.1109/CVPR.2014.81.
- [41] Matlab, “Train Object Detector Using R-CNN Deep Learning.” <https://de.mathworks.com/help/vision/ug/object-detection-using-deep-learning.html>.

APPENDICES

APPENDIX A: SUMMARY OF SUBJECT VIDEO AND MEDICAL SPECIALIST REVIEW	44
APPENDIX B: CALCULATION METHOD FOR THE MAGNITUDE OF FLUCTUATION.....	50

APPENDIX A: SUMMARY OF SUBJECT VIDEO AND MEDICAL SPECIALIST REVIEW

Table A1 Summary of subject videos.

Video No.	State of disease	Gender	Ages	Total duration (s)
1	Meniere's disease	Female	44	21
2	Meniere's disease	Female	60	28
3	Meniere's disease	Male	65	23
4	Meniere's disease	Male	81	29
5	Benign Paroxysmal Positional Vertigo	Female	80	18
6	Benign Paroxysmal Positional Vertigo	Female	75	29
7	Benign Paroxysmal Positional Vertigo	Female	80	29
8	Benign Paroxysmal Positional Vertigo	Male	73	29
9	Benign Paroxysmal Positional Vertigo	Female	80	29
10	Benign Paroxysmal Positional Vertigo	Male	73	29
11	Right vestibular disorder	Female	31	29
12	Right semicircular canal paralysis	Female	54	17
13	Meniere's disease	Male	37	19
14	Meniere's disease	Male	37	18
15	Left vestibular neuritis	Male	59	29
16	Left vestibular disorder	Male	55	27
17	Right sudden deafness	Male	47	29
18	Left Anterior Inferior Cerebellar Artery	Female	64	29
19	Right hunt	Female	54	19
20	Left vascular nerve compression	Female	68	18
21	Right hunt	Female	70	20
22	Right vestibular neuritis	Male	67	12

23	Right Meniere	Male	67	26
24	Right hunt	Male	80	21
25	Left vestibular neuritis	Male	71	11
26	Meniere's disease	Female	47	29
27	Meniere's disease	Female	72	29
28	Meniere's disease	Male	48	29
29	Meniere's disease	Male	57	8
30	Meniere's disease	Male	36	29
31	Meniere's disease	Male	65	29
32	Medulla oblongata bleeding	Male	44	13
33	Cerebellar disease Nystagmus for lower eyelid	Female	41	14
34	Neurovascular compression syndrome	Male	37	15
35	Spinocerebellar degeneration	Male	65	17
36	Congenital nystagmus	Male	28	7
37	Multiple system atrophy	Male	57	18

Table A2 Summary of medical specialist review.

Video No.	Review from medical specialist A, B, and C
1	A. This video is the nystagmus of a patient in the intermittent phase of Meniere's disease. The patient can open her eyes sufficiently. Therefore, the eyelid gap is wide, and the iris is well captured. The system captures the horizontal slow-phase component of the nystagmus.
	B. This video is a nystagmus finding in the interictal phase. The patient was able to open her eyes, and almost all of the iris is shown. A rapid eye movement in the right horizontal direction and a slow phase is recorded in the measurement waveform reproducing the actual nystagmus findings. Vertical nystagmus was not observed in the video, and the slow phase undetected in the measurement waveform. Therefore, the measurement waveform can reproduce the actual nystagmus.
	C. The system captures the horizontal nystagmus. Although the two slow phase velocities in the nystagmus are unstable, the system can be evaluated as generally measuring them without problems.
2	A. The video shows the nystagmus of a patient in the acute phase of Meniere's disease. The patient is in the acute phase of a vertiginous attack and may have difficulty opening her eyes sufficiently. As a result, the

eyelid gap is narrow, and it is usually difficult to capture the iris. However, the system captures a slow phase component of the nystagmus in the horizontal and vertical directions.

- B. The video shows the nystagmus of a patient in the acute phase of Meniere's disease. The patient seems to be difficult to open her eyes sufficiently. As a result, the eyelid gap is relatively narrow, and it is usually difficult to capture the iris completely. However, the waveform of this system captures the horizontal and vertical slow phase components of the nystagmus. Therefore, this system can be used even the eyelid gap is narrow.
- C. The vertical component of the nystagmus is well captured in the system. On the other hand, the horizontal part was lacking, resulting in some confusion in the results. This case is also in the acute stage of vertigo, and the nystagmus components may include various directions. Therefore, further analysis of the rotation component may help us to detect the disease more clearly.

A. The video shows the nystagmus of a patient in the acute phase of Meniere's disease. The patient is in the acute phase of a vertiginous attack and may have difficulty opening his eyes sufficiently. As a result, the eyelid gap is narrow, and it is usually difficult to capture the iris. However, the system can capture horizontal and vertical slow phase components of nystagmus during head-turning.

- 4
- B. The video shows the nystagmus of left Meniere's disease in the paroxysmal period. The entire iris is well captured in the image. The nystagmus is mainly in the right horizontal direction and has a slight rotation component in the image. The rapid phase and slow phase components of the right direction are evaluated in the measurement waveform. Vertical eye movements were not shown in any waveforms suggestive of nystagmus.
 - C. Although the eye movements could not be captured in the second half, the horizontal component was accurately captured in the first half. This is a case where the goggles used for recording need to be improved, and this analysis software is commendable.

A. The video shows the nystagmus of a patient with benign paroxysmal positional vertigo. The patient is in the acute phase of a vertiginous attack and may have difficulty opening her eyes sufficiently. As a result, the eyelid gap is narrow, and it is usually difficult to capture the iris. However, this system captures the horizontal slow phase component of the nystagmus.

- 9
- B. The video shows head-on nystagmus of BPPV. The nystagmus is mainly in the right horizontal direction and a slight rotation component in the image. All of the iris is unobserved in many cases, and the iris is unobserved in the eyelid gap in about 1/3 of the video. However, rapid and slow phase components in the right direction are evaluated in the measurement waveform. The rapid phase, which is presumably downward due to the gyration component, is reproduced in the vertical direction.
 - C. Although the frequency of nystagmus resolution is high in this case, the nystagmus is accurately captured in the horizontal component. Although some of the rapid phases are not fully grasped, the waveform can be evaluated as nystagmus in patients with vertigo, especially BPPV.
-

11	<p>A. The video shows the nystagmus of a patient with a vestibular disorder. The patient has difficulty opening her eyes sufficiently. As a result, the eyelid gap is narrow, and it is usually difficult to capture the iris. However, this system captures the horizontal slow-phase component of the nystagmus.</p> <p>B. Leftward nystagmus on the healthy side due to right vestibular dysfunction was observed. Although contact lenses were worn by the patient, most of the iris were visible. The left horizontal rapid-phase and slow-phase components are evaluated in this measurement waveform. The presence of contact lenses does not affect the analysis.</p> <p>C. Although nystagmus is difficult to evaluate with the naked eye due to its low frequency, this analysis confirms a horizontal component. The absence of a vertical element makes it possible to evaluate the nystagmus as an HC-BPPV.</p>
14	<p>A. The video shows the nystagmus of a patient with Meniere's disease in the intermittent phase. The patient has difficulty opening his eyes sufficiently. As a result, the eyelid gap is narrow, and it is usually difficult to capture the iris. However, this system captures the horizontal slow-phase component of the nystagmus.</p> <p>B. The eye movements are mainly left horizontal nystagmus, but there is a left downward oblique movement once every few strikes. In the measurement waveform, the slow phase in the left horizontal direction is visible. The vertical analysis also shows the waveform suggesting downward eye movement due to obliquity. Subtle vertical eye movements were captured.</p> <p>C. The nystagmus component in Meniere's disease is often complex. Even in cases such as the present case, where the nystagmus appears to have only a horizontal component to the naked eye, the analysis suggests that a vertical component is also present. The results of this analysis are sufficient for the analysis of nystagmus and may have clinical application.</p>
28	<p>A. The video shows the nystagmus of a patient with Meniere's disease in the intermittent phase. The patient can open his eyes sufficiently. Therefore, the eyelid gap is wide, and the iris is well captured. The system captures the horizontal slow-phase component of the nystagmus.</p> <p>B. The patient with Meniere's disease was in the interictal phase. The patient was able to maintain normal eye-opening. The left horizontal nystagmus is observed in the video. The left rapid-phase and slow-phase components in the horizontal movement are evaluated in the measurement waveform. This system captures nystagmus with low frequency in the intermittent phase.</p> <p>C. The video shows an example of what might be mistaken for impulsive eye movements by the naked eye because of the small amplitude of the nystagmus. However, the analysis captures horizontal nystagmus. The fact that the absence of a vertical component can be confirmed is also commendable.</p>
31	<p>A. The video shows the nystagmus of a patient with Meniere's disease in the intermittent phase. The patient can open his eyes sufficiently. Therefore, the eyelid gap is wide, and the iris is well captured. The system captures horizontal and vertical slow-phase components of nystagmus.</p> <p>B. The amplitude of the nystagmus is low, and the blink frequency is high even with the eye movement images because the patient with Meniere's</p>

disease is in the intermittent phase. Therefore, it is not easy to grasp eye movements. Nevertheless, the measurement waveform shows the rapid and slow phase components in the left horizontal direction. On the other hand, the vertical measurement shows nystagmus-like waveforms with rapid-phase and slow-phase components in the upper eyelid direction. However, it is difficult to identify them in the actual eye movement images.

- C. It is difficult to differentiate between peripheral and central nystagmus at first glance, as this case has both large and small amplitude components. The patient also had a brain tumor, and the presence of vertical nystagmus may provide clinically useful information, which is commendable.
-

A. The video shows the nystagmus of a patient with a Medulla oblongata bleeding. The patient is in the acute phase of a vertiginous attack and may have difficulty opening his eyes sufficiently. As a result, the eyelid gap is narrow, and it is usually difficult to capture the iris. However, this system captures horizontal and vertical slow-phase components of the nystagmus.

B. The eye movement images show that the eye is displaced to the right and that it is difficult to capture the entire iris due to the narrow eyelid gap. Nevertheless, leftward nystagmus observed frequently can be seen.

32 Although it lacks continuity in some places, the measurement waveform shows a rapid phase and a slow phase in the left horizontal direction.

- C. Although the frequency and amplitude of the nystagmus were considerable, the rapid phase of the horizontal component was not captured, which shows that the accuracy of the evaluation of the rapid phase is limited. However, it is sufficient to evaluate the slow phase. The fact that the vertical component is also captured is commendable. The fact that the vertical component also does not capture the rapid phase seems to be due to the limitation of the capturing capability of the infrared camera. Therefore, it is desirable to use a more powerful camera to capture the rapid phase more clearly.
-

A. The video shows the nystagmus of a patient with cerebellar disease nystagmus for the lower eyelid. The patient can open her eyes sufficiently. Therefore, the eyelid gap is wide, and the iris is well captured. The system captures the vertical slow-phase components of nystagmus.

B. The entire iris is captured in the second half of the recording, and rhythmic downward eye movement can be confirmed in the eye movement images. In the measurement waveform, the rapid downward and slow phase are evaluated in the second half of the images. There is a scene where the eyeball is significantly displaced to the right in the first half of the images. In such a situation in which the iris is partially missing, the measurement waveform does not reproduce the nystagmus.

- 33 C. The patient came to our hospital with a complaint of balance disorder due to spinocerebellar degeneration. The downward nystagmus was accurately captured, and the presence of a weak horizontal component could be confirmed. The fact that the nystagmus can be recognized even when the eyelid is lowered and half of the iris cannot be captured is commendable.
-

-
- 35
- A. The video shows the nystagmus of a patient with Spinocerebellar degeneration. The patient has difficulty opening his eyes sufficiently. As a result, the eyelid gap is narrow, and it is usually difficult to capture the iris. However, this system captures the vertical slow-phase component of the nystagmus.
 - B. Although about 1/3 of the iris is blocked by the upper eyelid in the eye movement images, the downward nystagmus can be recognized. Some oblique eye movements are included in the images. The waveform shows a rapid phase and a slow phase in the vertically downward direction. The waveform captures the nystagmus even if the entire iris is not recorded. A rightward movement due to the actual oblique movement is observed in the horizontal analysis. However, the rightward movement cannot be evaluated as a clear rapid-slow phase.
 - C. The patient came to our hospital with a complaint of balance disorder due to spinocerebellar degeneration. The downward nystagmus was accurately captured, and the presence of a weak horizontal component could be confirmed. The fact that the nystagmus can be recognized even when the eyelid is lowered and half of the iris cannot be captured, commendable.
-
- 37
- A. The video shows the nystagmus of a patient with multiple system atrophy. The patient has difficulty opening his eyes sufficiently. As a result, the eyelid gap is narrow, and it is usually difficult to capture the iris. However, this system captures horizontal and vertical slow-phase components of the nystagmus.
 - B. The nystagmus is predominantly downward and oblique with a suitable horizontal component in the eye movement images. Although the entire iris was not visible in some areas, the measurement waveform reproduced both horizontal and vertical eye movements.
 - C. The video shows a case of multiple system atrophy and central vertigo. Vertical nystagmus is the predominant finding. It can be seen that there is also a horizontal component. Clinically, the results of the analysis are consistent.
-

APPENDIX B: CALCULATION METHOD FOR THE MAGNITUDE OF FLUCTUATION

The fluctuation of $r(t)$ is calculated as follows: First, the absolute difference between $r(t)$ and $r(t+1)$ is calculated, using

$$rdiff(t) = |r(t) - r(t+1)|, \quad (B.1)$$

where $rdiff(t)$ is the absolute difference. Then, $rdiff(t)$ is categorized into false detection and true detection categories. The category is divided based on the value of $rdiff(t)$, which represents the variation of $r(t)$. If $rdiff(t)$ is larger than 10 pixels, then the detection is categorized as false detection. Otherwise, $rdiff(t)$ is categorized as true detection.

The average of false detection occurrence is defined as $rdiff_{false}$, which is calculated using:

$$rdiff_{false} = \frac{1}{T} \sum_{t=1}^T occ_{false}(t), \quad (B.2)$$

in which,

$$occ_{false}(t) = \begin{cases} 10, & rdiff(t) > 10 \\ 0 & \end{cases}, \quad (B.3)$$

where $occ_{false}(t)$ is the false detection occurrence.

The true detection average is calculated based on the total occurrence of $rdiff(t)$ which is lower than 10 pixels. The total occurrence of true detections, defined as $countrdiff_{true}$, is calculated using:

$$countrdiff_{true} = \sum_{t=1}^T occ_{true}(t), \quad (B.4)$$

where

$$occ_{true}(t) = \begin{cases} 1, & rdiff(t) \leq 10 \\ 0 & \end{cases}, \quad (B.5)$$

in which $occ_{true}(t)$ is a variable that takes a binary value, representing the occurrence of a true detection by 1; otherwise, it is 0. Then, the total value of $rdiff(t)$ for these true detections is defined as $totaldiff_{true}$, which is also calculated using:

$$totaldiff_{true} = \sum_{t=1}^T value_{true}(t), \quad (B.6)$$

in which,

$$value_{true}(t) = \begin{cases} rdiff(t), & rdiff(t) \leq 10 \\ 0 & \end{cases}, \quad (B.7)$$

where $value_{true}(t)$ is a variable that summarizes the value of true detection from $rdiff(t)$.

Finally, $rdiff_{true}$, which defines the true detection average, is calculated using

$$rdiff_{true} = \frac{totaldiff_{true}}{countdiff_{true}}. \quad (B.8)$$

Based on Equations (B.2) and (B.8), the fluctuation of $r(t)$ is calculated based on $rdiff_{false}$ and $rdiff_{true}$, using:

$$rdiff_{all} = rdiff_{true} + rdiff_{false}, \quad (B.9)$$

where $rdiff_{all}$ is the fluctuation of $r(t)$. The optimum q is defined as that which results in the minimum $rdiff_{all}$. Finally, all of the $x_0(t)$, $y_0(t)$, and $r(t)$ values from the optimum q are collected as the final detection result.