tvst

Article

Automatic Recording of the Target Location During Smooth Pursuit Eye Movement Testing Using Video-Oculography and Deep Learning-Based Object Detection

Masakazu Hirota[1,2], Takao Hayashi[1,2], Emiko Watanabe[2], Yuji Inoue[2], and Atsushi Mizota[2]

[1] Department of Orthoptics, Faculty of Medical Technology, Teikyo University, Itabashi, Tokyo, Japan
[2] Department of Ophthalmology, School of Medicine, Teikyo University, Itabashi, Tokyo, Japan

Correspondence: Masakazu Hirota, 2-11-1 Kaga, Itabashi, Tokyo 173-8605, Japan. e-mail: hirota.ortho@med.teikyo-u.ac.jp

Received: November 13, 2020
Accepted: March 28, 2021
Published: May 3, 2021

Keywords: eye movements; pursuit eye movements; video oculography; artificial intelligence; object detection

Citation: Hirota M, Hayashi T, Watanabe E, Inoue Y, Mizota A. Automatic recording of the target location during smooth pursuit eye movement testing using video-oculography and deep learning-based object detection. Transl Vis Sci Technol. 2021;10(6):1, https://doi.org/10.1167/tvst.10.6.1

Purpose: To accurately record the movements of a hand-held target together with the smooth pursuit eye movements (SPEMs) elicited with video-oculography (VOG) combined with deep learning-based object detection using a single-shot multibox detector (SSD).

Methods: The SPEMs of 11 healthy volunteers (21.3 ± 0.9 years) were recorded using VOG. The subjects fixated on a moving target that was manually moved at a distance of 1 m by the examiner. An automatic recording system was developed using SSD to predict the type and location of objects in a single image. The 400 images that were taken of one subject using a VOG scene camera were distributed into 2 groups (300 and 100) for training and validation. The testing data included 1100 images of all subjects (100 images/subject). The method achieved 75% average precision (AP_{75}) for the relationship between the location of the fixated target (as calculated by SSD) and the position of each eye (as recorded by VOG).

Results: The AP_{75} for all subjects was 99.7% ± 0.6%. The horizontal and vertical target locations were significantly and positively correlated with each eye position in the horizontal and vertical directions (adjusted $R^2 \geq 0.955$, $P < 0.001$).

Conclusions: The addition of SSD-driven recording of hand-held target positions with VOG allows for quantitative assessment of SPEMs following a target during an SPEM test.

Translational Relevance: The combined methods of VOG and SSD can be used to detect SPEMs with greater accuracy, which can improve the outcome of clinical evaluations.

Introduction

Eye movements include the ability to fixate and track visual stimuli. In most ophthalmology clinics, the examiner evaluates smooth pursuit eye movements (SPEMs) by subjectively noting their accuracy in relation to a target that is being moved manually by an examiner while the patient follows it with his or her eyes.[1–8] In clinical settings, eye movement tests are not usually recorded, even though such recordings could be used to evaluate changes in eye function over time, which would be particularly useful for difficult cases.

In the laboratory, several methods can be used to quantify eye movements, including the search coil method,[9,10] a electrooculography,[11] and video-oculography (VOG).[10] Each of these techniques can record the eye movements of both eyes simultaneously while the subjects are fixating their focus on a moving target. However, laboratory methods have not been introduced in clinical practice for two reasons. First, the type of target that is used at the laboratory level is not the same as the target that is available for use in the clinical setting. To achieve accuracy at the laboratory level, it is necessary to present to the subjects a predetermined target according to the programming codes. In the clinical setting, the examiner modifies the movement of the target as appropriate to examine the suspected abnormality. For children, the examiner uses a target, such as an anime or game character, to hold their attention.[12–16] Second, VOG has an intrinsic target display, which does not allow for flexibility of target movement during the examination as might be desired depending upon the eye movement disorder.

We hypothesized that a combination of VOG and a deep learning-based object detection algorithm allow for laboratory methods to be extended to the clinical practice. Deep learning-based object detection technology can predict the location and types of objects in one image.[17–19] Furthermore, the algorithm for deep learning-based object detection can detect objects in real-time using a movie with a processing speed of > 30 frames per second (fps). This processing speed is faster than that of simpler conventional algorithms that use raster scans per image.[20] Therefore, we hypothesized that by using VOG to record patient SPEMs and using deep learning-based object detection to record target movements, the combination system simultaneously measures SPEM and target without requiring significant changes to the clinical examination procedure.

Thus, this study aimed to determine whether a deep learning-based object detection technique could be used to quantify hand-held target movements, which could then be combined with the standard VOG method to more accurately assess SPEM. We evaluated the target location in relation to the eye positions of healthy volunteers. The experiments were designed to mimic the ophthalmological clinical setting, although the clinical technique was slightly modified.

Methods

Subjects

A total of 11 volunteers (age = 21.3 ± 0.9 years [mean ± standard deviation]) participated in this study. All subjects underwent complete ophthalmologic examinations, including determination of the ocular dominance using the hole-in-the-card test, best-corrected visual acuity at a distance (5.0 m), near point of convergence, stereoscopic acuity at 40 cm (Titmus Stereotest; Stereo Optical Co., Inc., Chicago, IL, USA), heterophoria by the alternating cover test at near (33 cm) and at distance (5.0 m) assessments, and fundus examinations. Stereoacuity was converted into the logarithm of the arc second (log arcsec). Participants were excluded if they had a refractive error of ≥ 10.0 diopters (D).

The demographics of the subjects are presented in Table 1. The mean ± standard deviation of the refractive errors (spherical equivalents) of the dominant eye was −2.95 ± 2.46 D and that of the nondominant eye was −2.70 ± 2.61 D. The best-corrected visual acuity was 0.0 logMAR units or better in all subjects. The average heterophoria was −1.4 ± 0.9 prism diopters (PD) at distance and −3.2 ± 4.2 PD at near. All healthy volunteers had a stereo acuity of 1.60 log arcsec (40 seconds).

Informed consent was obtained from all subjects after the nature of the study, and possible complications were explained to them. This investigation adhered to the tenets of the World Medical Association Declaration of Helsinki. Furthermore, informed

Table 1. Demographics of the Subjects

		SE (D)		Angle of Deviation (PD)		
Subject ID	Age (y)	Dominant Eye	Nondominant Eye	Near	Far	Stereo Acuity (Log Arcsec)
S1	21	−0.88	−0.38	−4	−2	1.60
S2	24	−7.13	−6.88	−4	−2	1.60
S3	21	−0.63	−0.13	−8	−2	1.60
S4	21	−2.50	−2.88	−10	−2	1.60
S5	21	−8.13	−8.00	−2	0	1.60
S6	21	−2.38	−2.38	+2	−1	1.60
S7	21	0.00	+0.25	+6	0	1.60
S8	21	−3.88	−4.00	−6	−2	1.60
S9	21	−3.25	−3.38	−1	0	1.60
S10	22	−2.25	−0.61	−4	−2	1.60
S11	21	−1.50	−1.38	−4	−2	1.60
Mean	21.4	−2.95	−2.70	−3.2	−1.4	1.60
SD	0.9	2.46	2.61	4.2	0.9	0.00

Minus and plus signs in the angle of deviation indicate exodeviation and esodeviation of phoria, respectively. Stereo acuity of 1.60 log arcsec is equal to 40 seconds.

S, subject; SE, spherical equivalent; D, diopter; PD, prism diopter; log arcsec, logarithm of arc second; SD, standard deviation.

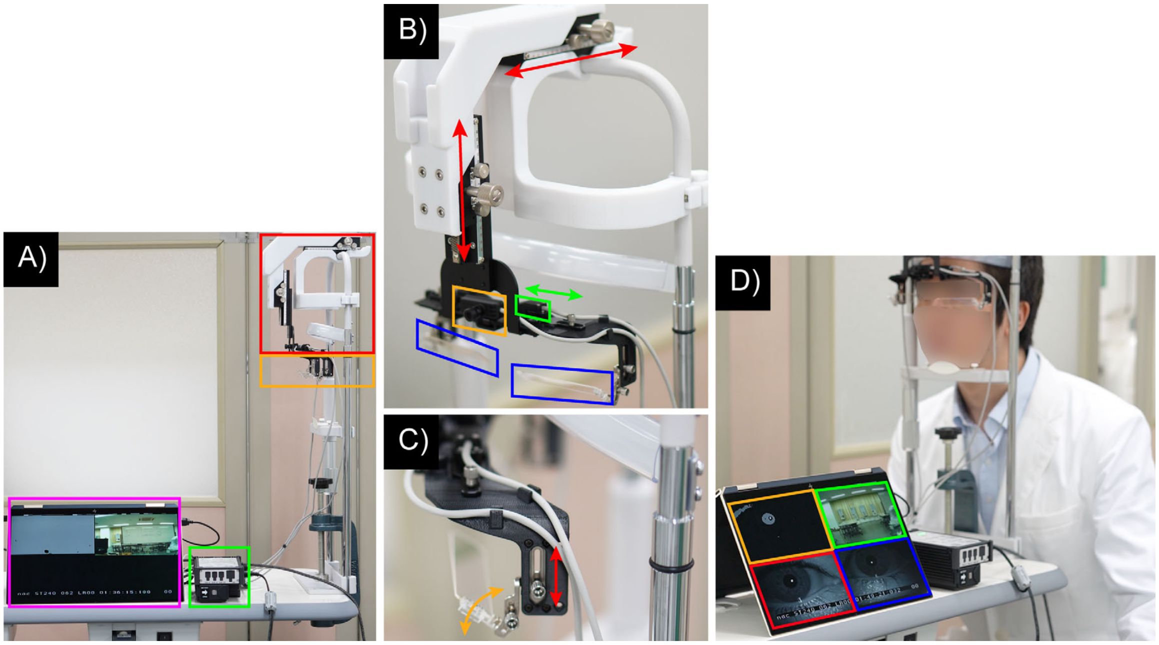

Figure 1. Video oculography. The exterior of the VOG (**A**). The total area of the red and orange squares indicates the VOG apparatus, whereas the red and orange squares indicate the zoomed in images of (**B**) and (**C**), as described below. The green square indicates the controller of the VOG. The purple square indicates the output image. In **B**, the orange, green, and blue squares indicate the scene camera, eye camera (there is also one for the right eye on the other side), and half-mirrors, respectively. The scene camera can rotate 60 degrees in pitch. Both eye cameras can horizontally shift to 1.3 cm (total = 2.6 cm) to adjust the pupillary distance (green two-direction arrow line). The base position of the VOG can adjust 8.0 cm horizontally and vertically (red two-direction arrow lines). In **C**, the half-mirror can shift 2.5 cm vertically (red two-direction arrow line) and rotate 30 degrees in pitch (orange two-direction arrow line). Scenes from the eye movement recording are shown in (**D**). The red and blue squares indicate the images of the right and left eyes, respectively, which were obtained by the eye cameras. The orange square indicates the binary image in the right eye used to confirm eye tracking accuracy during examination. The green square indicates the images that merged the gaze of both eyes to the real scenes that were recorded by a scene camera with a sampling rate of 29.97 Hz and a delay of ≤ 52 ms. VOG, video oculography.

consent was obtained for the publication of identifying information/images that were sourced from an online open-access publication. The Institutional Review Board of Teikyo University approved the experimental protocol and consent procedures (approval no. 18–161).

Apparatus

Eye movements while tracking the target were recorded using a commercial VOG (EMR-9, NAC Image Technology Inc., Tokyo, Japan; Fig. 1). The VOG device determined the eye positions by detecting the corneal reflex and pupil center that were created by the reflection of a near-infrared light with a sampling rate of 240 Hz (green square in Fig. 1B). The measurement error was 0.2 degrees–0.5 degrees (interquartile range) at a distance of 1.0 m. The scene camera recorded the real scenes (resolution = 640 × 480 pixels; angle of view = ±31 degrees from the center of the scene camera) with a sampling rate of 29.97 Hz (orange square in Fig. 1B). The images obtained by the eye camera and scene camera were sent to the controller (green square in Fig. 1A). The controller computes the gaze position of both eyes from the corneal reflex and pupil center (red, blue, and orange squares in Fig. 1D); then, the gaze positions are merged with the real scenes at a delay of ≤ 52 ms (green square in Fig. 1D). VOG outputs the recorded eye position and pupillary diameter of each eye in a comma separated values file and the video recorded by the scene camera in an M4F file. In this study, the VOG device was placed on a chin stand that could be moved 8.0 cm horizontally and 8.0 cm vertically. In this study, to avoid blocking the subject's gaze during examination, the eye camera was placed at the top of the VOG device to capture the subject's eyes using a half-mirror. VOG integrates the positional relationship between the scene camera and the eye camera through gaze calibration.

The position of the subject's eyes was set at an eye-level marker. Then, the calibration plate was set at 1.0 m. Noise caused by reflections of the lens was noted in 2 of 11 participants, which was then avoided by manually adjusting the tilt of the half-mirror. The laser pointers were placed beside the eye-level marker and above the scene camera to align the position of the subject's eyes with the center of the scene camera. All subjects underwent a calibration test to adjust the position of their gaze on the images of the scene camera and under binocular conditions with fully corrected glasses before performing the eye movement test. During calibration, all subjects were asked to fixate nine red cross targets (visual angle = 0.1 degrees) on a white calibration plate. From one to nine, the nine red crosses of the targets were set at the following parameters: (horizontal of 0.0 degrees and vertical of 0.0 degrees), (0.0 degrees and 20.0 degrees), (20.0 degrees and 20.0 degrees), (20.0 degrees and 0.0 degrees), (20.0 degrees and −20.0 degrees), (0.0 degrees and −20.0 degrees), (−20.0 degrees and −20.0 degrees), (−20.0 degrees and 0.0 degrees), and (−20.0 degrees and 20.0 degrees), respectively. The center of the calibration plate was defined as 0 degrees; the right and upper halves of the screen were defined as the positive sides; and the left and lower halves were defined as the negative sides.

Procedures

Recording Eye Movements Following a Target on a Video Screen

The accuracy of the VOG method was evaluated in an ideal environment as a preliminary study. Two subjects (each aged 34 years) participated in the preliminary study. The target was a rabbit-like character (Fig. 2). The size of the target was 10 × 10 cm, which subtended a visual angle of 5.7 degrees at 1.0 m. The target was displayed on a 24-inch liquid crystal monitor. The center of the monitor was defined as 0 degrees; the right and upper halves of the monitor were defined as the positive sides; and the left and lower halves were defined as the negative sides. The target was moved ±10 degrees with a random velocity of ≤10 degrees/s, which was preset by a computer. The subjects were seated in a well-lit room (600 lx) wearing fully corrective spectacles. The subject's head was fixed with a chin and forehead rest. The subjects were asked to fixate their focus on the nose of the target, whose visual angle was 0.1 degrees at 1.0 m, for 60 seconds.

The target location was significantly correlated with both eye positions. The horizontal target locations were significantly and positively correlated with the horizontal dominant (adjusted $R^2 = 0.989$, $P < 0.001$) and nondominant (adjusted $R^2 = 0.989$, $P < 0.001$) eye positions (Figs. 3A, 3B). The vertical target locations were significantly and positively correlated with the vertical dominant (adjusted $R^2 = 0.987$, $P < 0.001$) and nondominant (adjusted $R^2 = 0.987$, $P < 0.001$) eye positions (Figs. 3C, 3D).

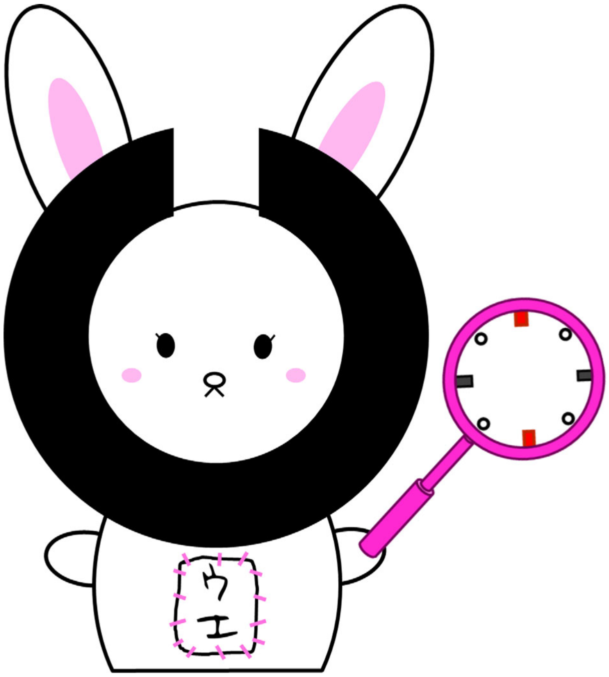

Figure 2. Examination target. The rabbit-like character is a mascot of the Department of Orthoptics, Teikyo University (created by Mika Suda). The subjects were asked to fixate on the nose of the target, whose visual angle was 0.1 degrees at 1.0 m during the eye movement tests.

Recording the Position of the Target When Moved by Hand

Eye Movement Test. The main study used the same target as that used in the preliminary study (see Fig. 2). The target was manually moved within ±15 degrees for 60 seconds by an examiner. All subjects were seated in a well-lit room (600 lx) wearing fully corrective spectacles. Each subject's head was stabilized with a chin rest and forehead rest. The subjects were asked to fixate on the nose of the target, whose visual angle was 0.1 degrees at 1.0 m, during the eye movement test.

Algorithm for Target Detection

The object detection algorithm was used for the single-shot multibox detector (SSD).[18] The SSD analyzed the input image using the base convolutional neural network (CNN) for extracting the feature map and the branch from the base CNN into two other CNNs to predict the object category and location. The SSD output surrounds the range of detected objects within a rectangle called a bounding box and displays the name of the object category and predicted value. In this study, a pretrained visual geometry

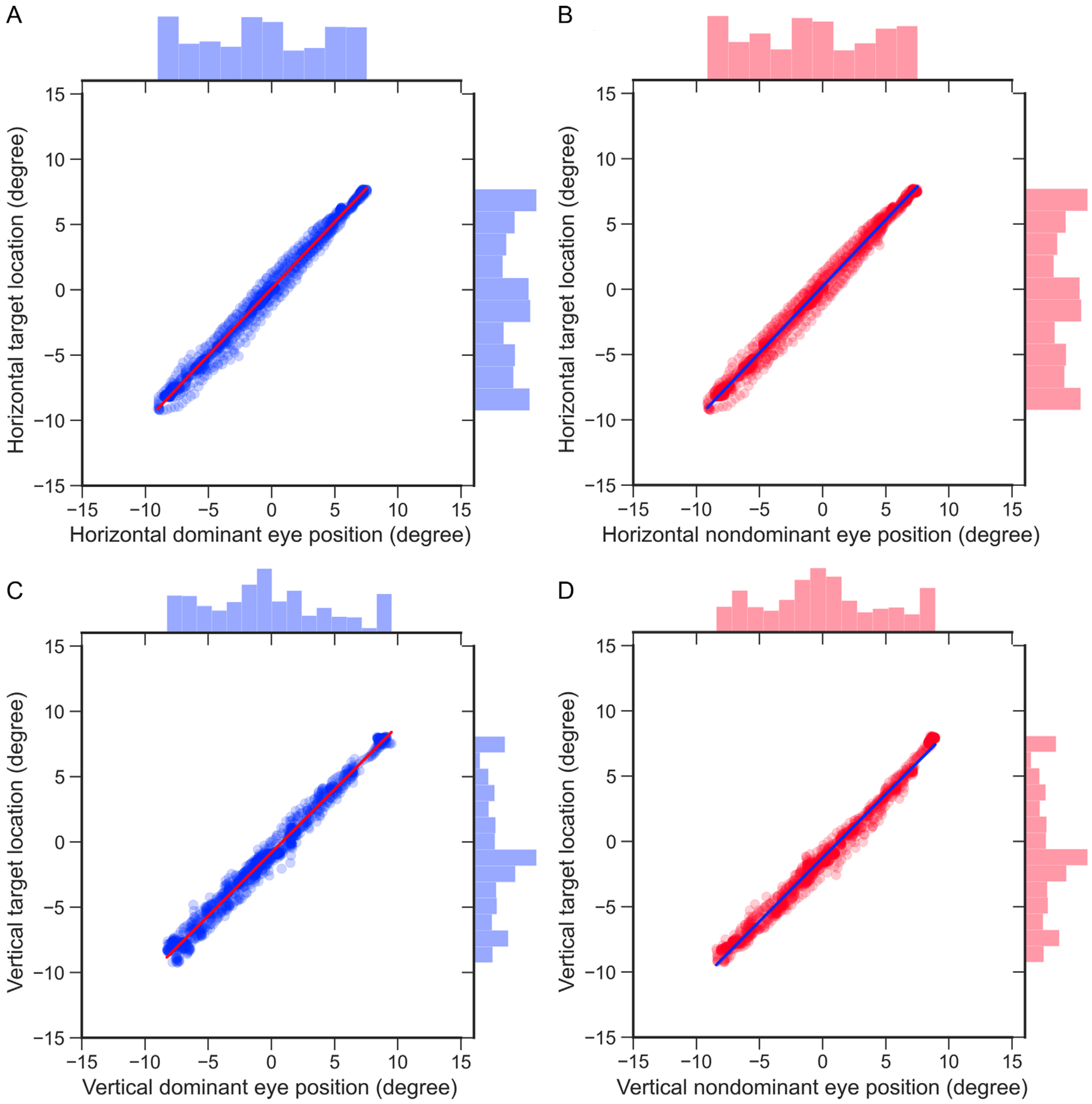

Figure 3. Correlations between horizontal (A, B) and vertical (C, D) target locations and eye positions in a preliminary study (*n* = 2). The blue and red dots indicate the relationships between the target location and the dominant or nondominant eye position of two subjects. The red and blue lines indicate regression lines. Histograms at the upper and right sides indicate the distribution between the target location and the dominant or nondominant eye position. Regression equation of **A**: 0.003 + 1.021 times. Regression equation of **B**: 0.003 + 1.019 times. Regression equation of **C**: 0.004 + 0.968 times. Regression equation of **D**: 0.004 + 0.966 times.

group-16 (VGG16) was used as the base CNN.[21,22] The object category was set to two: target and nontarget (background).

A total of 500 images were extracted from the video recordings of the eye movement test that was performed on subject 1, which had been recorded by a scene camera. The area of the target was annotated using LabelImg (Tzutalin). The area of the target was defined to the edge of the target image, and annotation was performed by an examiner (author M.H.). Then, the training, validation, and test datasets for subject 1 were randomly divided into 300 (60%), 100 (20%), and 100 (20%) images, respectively. Moreover, 100 images from subjects 2 through 11 were extracted and annotated as test data.

The target images were resized to 300 × 300 pixels in the preprocessing; then, data augmentation was applied to the target images. In the training phase, we

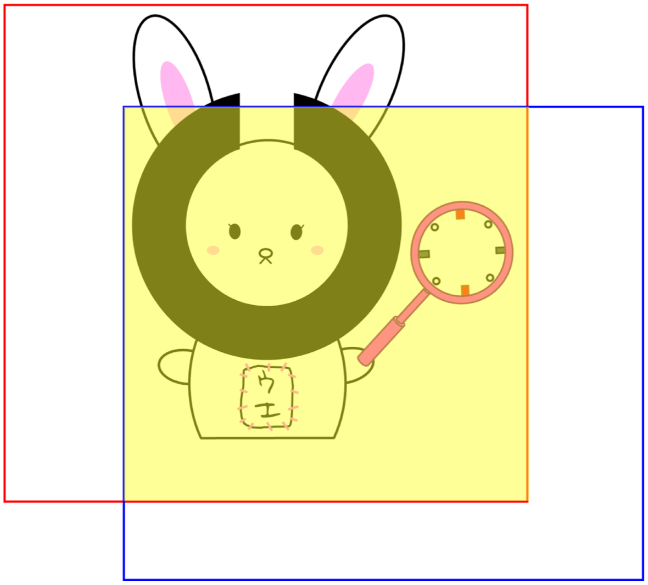

Figure 4. **Intersection of union.** The blue and red squares indicate the predicted bounding box and ground-truth bounding box, respectively. The yellow area is the overlapping region of the predicted bounding box and ground-truth bounding box. The intersection of union was calculated as follows: yellow area/(blue square area + red square area).

set the following parameters: 100 epochs, batch size of 32, and an Adam optimizer with a learning rate of 0.001. The validation was performed every 10 epochs.

The SSD model was evaluated by the average precision of the test dataset. The intersection of union (IoU) was calculated by dividing the area of overlap between the predicted bounding box and the ground-truth bounding box, which included the target name and target location as manually defined by the examiner (author M.H.), as well as the area of union of both bounding boxes (Fig. 4). We defined an IoU of $\geq$ 75% as "correct" (AP_{75}). Subsequently, the results of the test dataset were classified as follows: true positive (TP), the predicted bounding box has been covered with a ground-truth bounding box (IoU $\geq$ 75%); false positive (FP), the predicted bounding box has been covered with a ground-truth bounding box (IoU < 75% and IoU $\neq$ 0); false negative (FN), the predicted bounding box has not been covered with a ground-truth bounding box (IoU = 0); and true negative, the predicted bounding box and ground-truth bounding box did not exist. Precision was defined as the percentage at which the IoU can be predicted correctly with an accuracy of $\geq$ 75%. This was calculated by TP/(TP + FP). Recall was defined as the percentage at which the ratio of the bounding box at apposition can be accurately predicted if the IoU is $\geq$ 75% and was calculated by TP/(TP + FN). The AP_{75} was calculated by the integral precision and recall.

We used Python version 3.6.5 on Windows 10 (Microsoft Co., Ltd., Redmond, WA, USA) with the following libraries: Matplotlib 3.3.2, Numpy 1.18.5, OpenCV 3.3.1, Pandas 1.1.3, Pytorch 1.6.0, Scikit-learn 0.23.2, and Seaborn 0.11.0.

Tracking Target Motion

The SSD method involves drawing the location of the object within a bounding box. The bounding box was computed from two coordinates of (X_{min} and Y_{min}) and (X_{max} and Y_{max}). The center of the object coordinates (Cx and Cy) was determined by ($[X_{max} + X_{min}]/2$, $[Y_{max} + Y_{min}]/2$). Thus, the target location was defined as the center of the bounding box, and a program output the target location in each frame findings to an Excel file (Microsoft Co., Ltd.) with the recording time synchronized with VOG in the inference phase.

Data Analyses

Data were excluded when the change in pupil diameter was > 2 mm/frame due to blinking,[23] and the percentage of missing values (0.4% ± 0.7% for all subjects) was replaced with a linearly interpolated value that was calculated from an algorithm written with Python 3.6.5. The horizontal and vertical eye movements were analyzed, and the SPEM and saccadic eye movements (SEMs) were identified using a velocity-threshold identification (I-VT) filter.[24] The I-VT filter was used to classify eye movements based on the velocity of the directional shifts of the eye. An SEM was defined as the median velocity of 3 consecutive windows > 100 degrees/second. Then, the eye position data at 240 Hz were synchronized with the target data at 29.97 Hz.

The peaks indicating the latency of the waveforms of the target locations and eye positions that occurred when the direction of movement was reversed were determined visually (Fig. 5). The latency of SPEM was calculated from the difference between the horizontal and vertical target location peaks and the dominant and nondominant eye position peaks. The latency of each eye was calculated at least three times and averaged for each subject.

Statistical Analyses

The inter-rater reliability between the target location and both eye positions was analyzed using a Bland-Altman plot.[25,26] A paired *t*-test and simple linear regression analysis were performed to check for fixed and proportional biases between the target location and both eye positions. The mean value of the

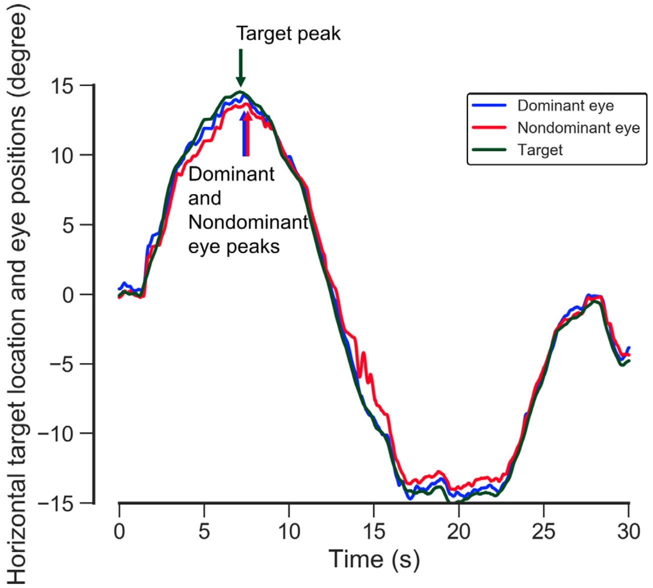

Figure 5. Graph showing the latencies of SPEM for the dominant and nondominant eyes of subject 1. The blue, red, and green lines indicate the horizontal dominant and nondominant eye positions and target location during the eye movement test. The latencies of SPEM were calculated from the difference between the target location peak and the dominant and nondominant eye position peaks.

difference between the target location and both eye positions indicated that there was a reliability of ≤ 0.5 degrees if there was no fixed or proportional bias because the measurement error of VOG was within 0.5 degrees. The reliability of the target location and eye position for both eyes was analyzed using intraclass correlation coefficients. The relationships between the target location and both eye positions were assessed using simple linear regression analysis.

To confirm the effect of an artifact of VOG and/or SSD (i.e. eye lashes or deformation of bounding box), the latency of SPEM was calculated. The differences between the latencies of the dominant eye and nondominant eye were analyzed with the paired *t*-tests. The relationship between latencies of horizontal and vertical SPEM within both eyes was assessed using simple linear regression analysis.

SPSS version 26 (IBM Corp., Armonk, NY, USA) was used to determine the significance of the differences, and a *P* value of < 0.05 was considered to be statistically significant.

Results

Consistency of VOG and SSD

SSD took 880 seconds for 100 epochs of training. A representative result (subject 7) for our SSD is shown in Movie 1 and Figure 6. SSD could correctly track a target that an examiner manually moved. SSD was analyzed the recording video with a mean speed of 0.031 image/second (about 32.25 fps). The AP_{75} for all subjects was 99.7% ± 0.6% (Table 2).

The horizontal target location (−0.03 degrees ± 0.81 degrees) did not differ from the horizontal dominant (−0.03 degrees ± 0.84 degrees, $P = 0.95$) or nondominant (0.04 degrees ± 0.84 degrees, $P = 0.71$) eye positions for any of the subjects. The vertical target location (0.15 degrees ± 1.15 degrees) did not differ from the horizontal dominant (0.08 degrees ± 1.22 degrees, $P = 0.79$) or nondominant (0.30 degrees ± 1.11 degrees, $P = 0.55$) eye positions for any of the subjects. The mean values of the differences between the horizontal target location and both eye positions were −0.01 degrees (95% limits of agreements [95% interunit reliability {LoA}], −0.87 to 0.85) in the dominant eye and 0.07 degrees (95% LoA = −1.04 to 1.19) in the nondominant eye, and the correlation between the horizontal target location and both eye positions was not significant (dominant eye = adjusted $R^2 < 0.000$, $P = 0.83$; nondominant eye = adjusted $R^2 < 0.000$, $P = 0.88$; Figs. 7A, 7B). The mean values of the differences between the vertical target location and both eye positions were −0.07 degrees (95% LoA = −1.51 to 1.38) in the dominant eye and 0.15 degrees (95% LoA = −1.33 to 1.64) in the nondominant eye, and the correlation between the horizontal target location and both eye positions was not significant (dominant eye = adjusted $R^2 < 0.000$, $P = 0.76$; nondominant eye = adjusted $R^2 < 0.000$, $P = 0.90$; Figs. 7C, 7D). The target location was significantly reliable for predicting the eye position in both eyes (the horizontal generalizability coeficient was 0.944, $P < 0.001$; the vertical generalizability coefficient was 0.904, $P < 0.001$).

The horizontal target location was significantly and positively correlated with the horizontal dominant (adjusted $R^2 = 0.984$, $P < 0.001$) and nondominant (adjusted $R^2 = 0.983$, $P < 0.001$) eye positions of all subjects (Figs. 8A, 8B). The vertical target location was significantly and positively correlated with the vertical dominant (adjusted $R^2 = 0.955$, $P < 0.001$) and nondominant (adjusted $R^2 = 0.964$, $P < 0.001$) eye positions of all subjects (Figs. 8C, 8D).

The latencies of the horizontal and vertical SPEM were 99.0 ± 25.6 and 117.0 ± 34.2 ms, respectively, for the dominant eye of all subjects ($P = 0.22$). The latencies of the horizontal and vertical SPEMs for the nondominant eye were 111.0 ± 37.0 and 126.0 ± 35.6 ms, respectively, for all subjects ($P = 0.34$). The latencies of horizontal SPEM were significantly and positively correlated with the latencies of vertical SPEM in both the dominant (adjusted $R^2 = 0.761$,

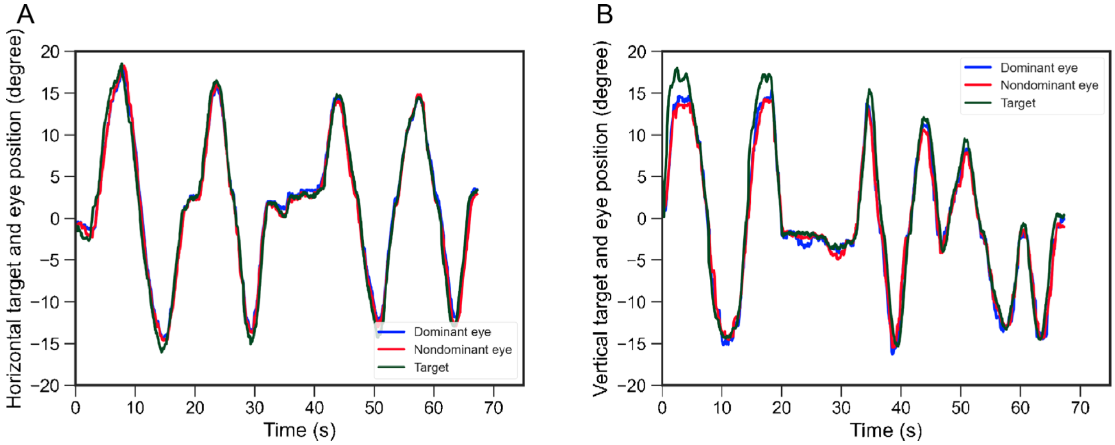

Figure 6. Horizontal (A) and vertical (B) target locations and eye positions when the examiner moved the target by hand in Movie 1. The blue, red, and green lines indicate the horizontal dominant and nondominant eye positions and target location during the eye movement test in subject 7.

Table 2. Results of the Average Precision and Latency

		Latencies (ms)			
		Horizontal		Vertical	
Subject ID	Average Precision	Dominant Eye	Nondominant Eye	Dominant Eye	Nondominant Eye
S1	97.7	66	66	99	99
S2	99.9	66	66	66	66
S3	99.2	132	99	132	132
S4	100.0	66	66	66	66
S5	100.0	99	132	132	132
S6	100.0	99	165	99	165
S7	100.0	99	165	99	165
S8	100.0	99	99	132	132
S9	100.0	132	132	165	132
S10	100.0	132	132	165	165
S11	100.0	99	99	132	132
Mean	99.7	99.0	111.0	117.0	126.0
SD	0.6	25.6	37.0	34.2	35.6

S, subject; SD, standard deviation.

$P < 0.001$) and nondominant (adjusted $R^2 = 0.765$, $P < 0.001$) eyes.

Additional Experiment for Patients With Strabismus

One patient (age = 38 years) with postsurgical congenital superior oblique muscle palsy participated in an additional experiment to test the scope of clinical applicability. This patient underwent complete ophthalmologic examinations, including a determination of ocular dominance using the hole-in-the-card test, best-corrected visual acuity at distance, the near point of convergence, stereoscopic acuity at 40 cm, heterotropia by alternate cover test at near and at distance, and fundus examinations. Stereoacuity was converted to log arcsec.

The dominant eye was the right eye, and the left eye had undergone surgery for strabismus 30 years ago. The patient was examined in the natural head position so that binocular vision could be maintained, because the patient had abnormal head positions: face turned to the right, head tilt to the right, and chin down. The vertical palpebral fissure was 4.0 mm in the dominant eye and 3.0 mm in the nondominant eye, respectively. The spherical equivalent of the dominant eye was −0.75 D and that of the nondominant eye was −0.75 D. The

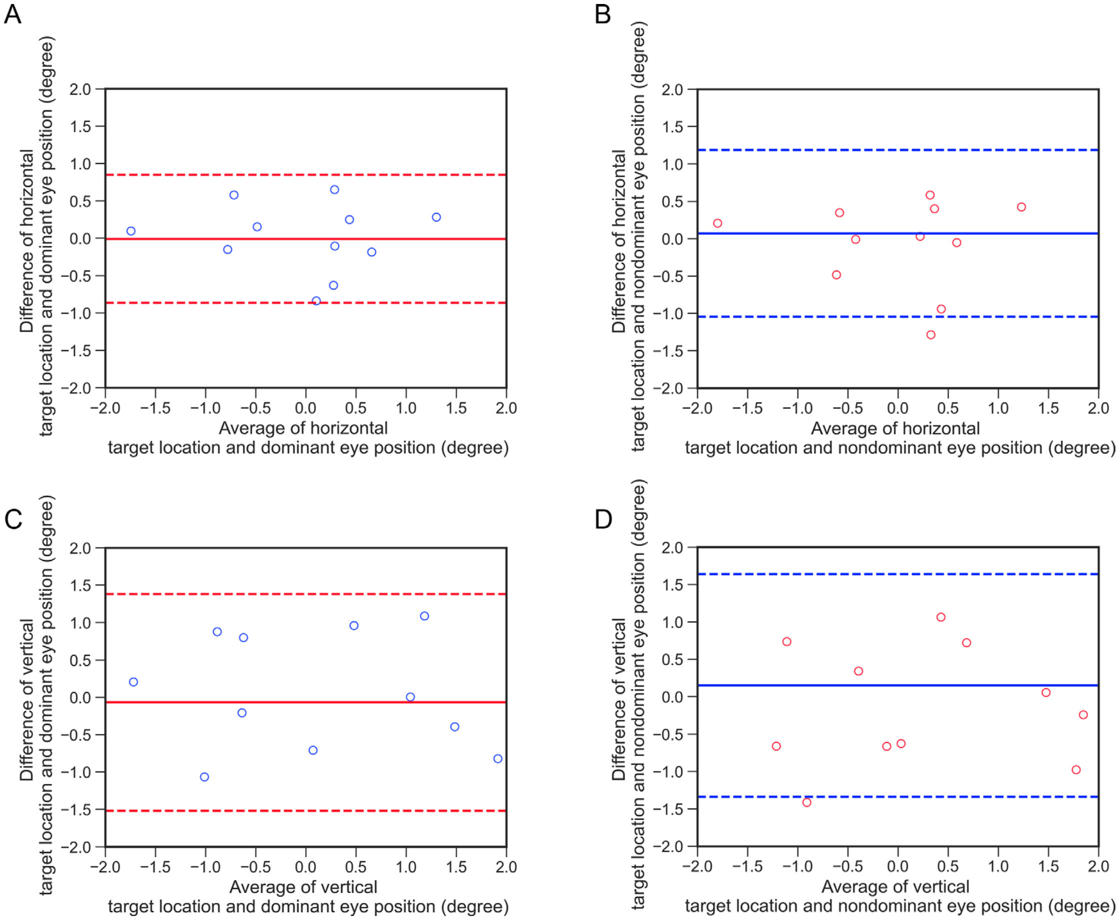

Figure 7. Interrater reliability between horizontal (A, B) and vertical (C, D) target locations and eye positions. The blue and red dots indicate the interrater reliability between the target location and the dominant or nondominant eye position. The red and blue solid lines show the mean value of the difference between the horizontal or vertical target location and the dominant or nondominant eye position. The red and blue dashed lines indicate 95% limits of agreement, which is defined as the mean ± 1.96 standard deviation.

best-corrected visual acuity was 0.0 logMAR in each eye. The horizontal and vertical heterotropia was 1.0 PD base-out and 1.0 PD base-up at distance and 4.0 PD base-in and 7.0 PD base-up at near. The stereo acuity was 1.60 log arcsec (40 seconds). The patient underwent the eye movement test for 45 seconds.

The results are shown in Figure 9. The horizontal SPEM in the nondominant eye was always delayed in relation to the movement of the horizontal target (see Fig. 9A). The vertical SPEM in the nondominant eye could not be assessed accurately due to ptosis (see Fig. 9B).

Discussion

VOG is rarely used in daily clinical practice because it has an intrinsic target display that does not allow for flexibility in the target movement during examination, as might be desired depending on the eye movement disorder. In this study, the deep learning-based object detection technique SSD was used to quantify the hand-held target movements, which was combined with the standard VOG method to assess SPEMs. The SSD method detected the target with high accuracy (Movie 1), and the target location was significantly and positively correlated with the VOG-recorded positions of both eyes (Figs. 6–8). The small variation in values in the preliminary study (Figs. 3, 8) may be attributable to the fact that the SSD bounding box was more susceptible to deformation; because the examiner moved the target by hand, the center coordinates of the bounding box were converted from pixels to degrees, which may have caused slight differences in the calculation of the coordinates. Nevertheless, the AP_{75} was 99.7% for all subjects. These findings suggest that deep learning techniques can be used to record the movements of hand-held stimulus targets so that the recorded movement of the stimulus can be

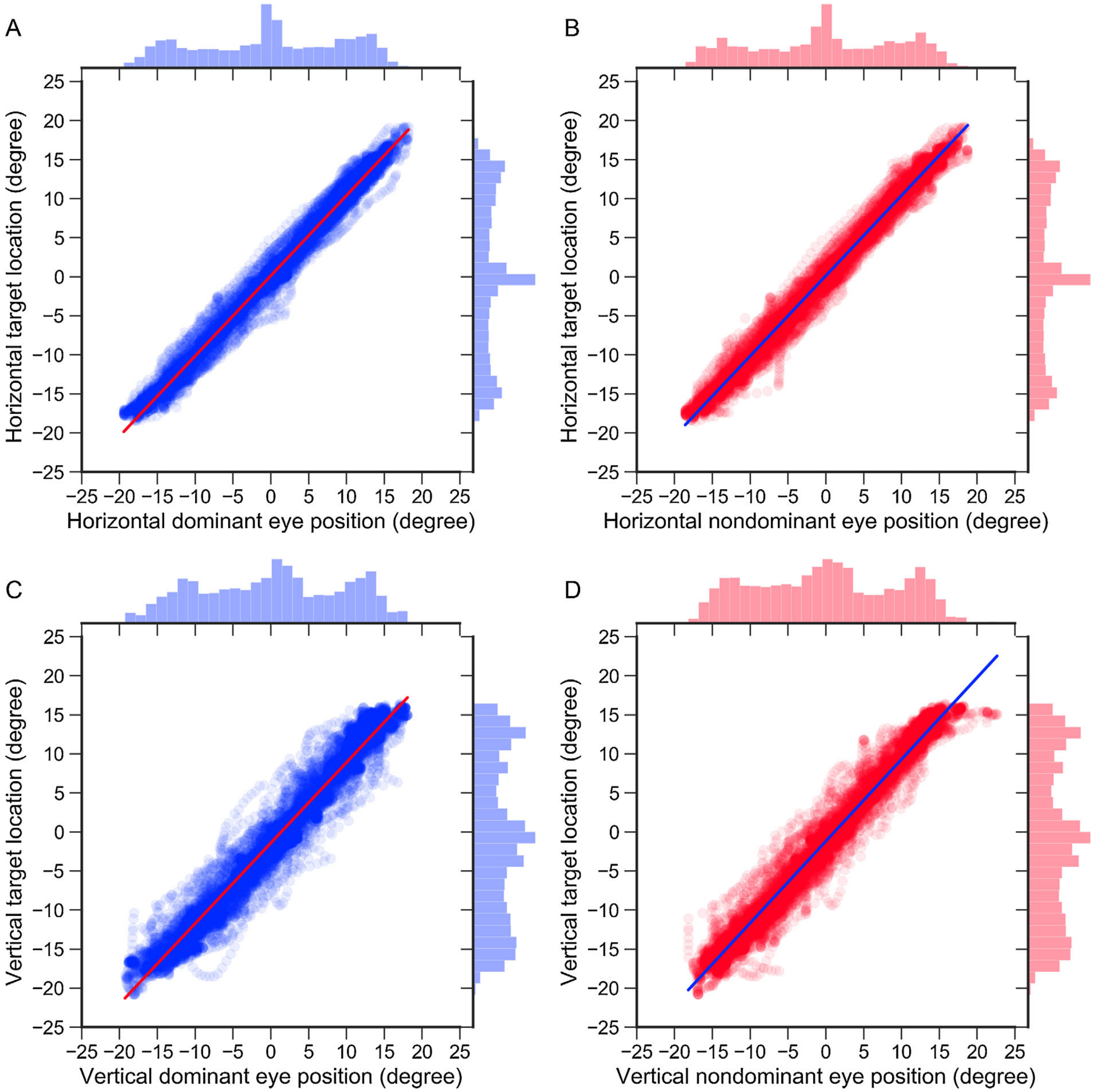

Figure 8. Correlations between horizontal (A, B) and vertical (C, D) target locations and eye positions of all subjects. The blue and red dots indicate the relationships between the target location and the dominant or nondominant eye position. The red and blue lines indicate regression lines. Histograms at the upper and right sides indicate the distribution between the target location and the dominant or nondominant eye position. Regression equation of **A**: 0.121 + 1.027 times. Regression equation of **B**: 0.255 + 1.021 times. Regression equation of **C**: −1.191 + 1.048 times. Regression equation of **D**: −1.421 + 1.031 times.

quantified accurately and thus can be compared more reliably with the eye movements that are recorded by VOG.

The mean latencies of the horizontal SPEMs in this study were consistent with the earlier studies performed by Erkelens and Engel.[27,28] In addition, our findings support those of Rottach et al., who reported that the latencies of horizontal and vertical SPEMs were not significantly different.[29] These findings suggest that the current system can determine SPEM in healthy individuals with lower interference from artifacts (i.e. eyelashes or deformation of the bounding box) in VOG and/or SSD.

We evaluated a patient who underwent surgery for strabismus, in order to show that our technique can detect clinically relevant SPEMs. The combination of VOG and SSD elucidated the delay of horizontal SPEM in the nondominant (affected) eye due to restoration of the binocular vision and compensatory head posture after surgery for strabismus (see

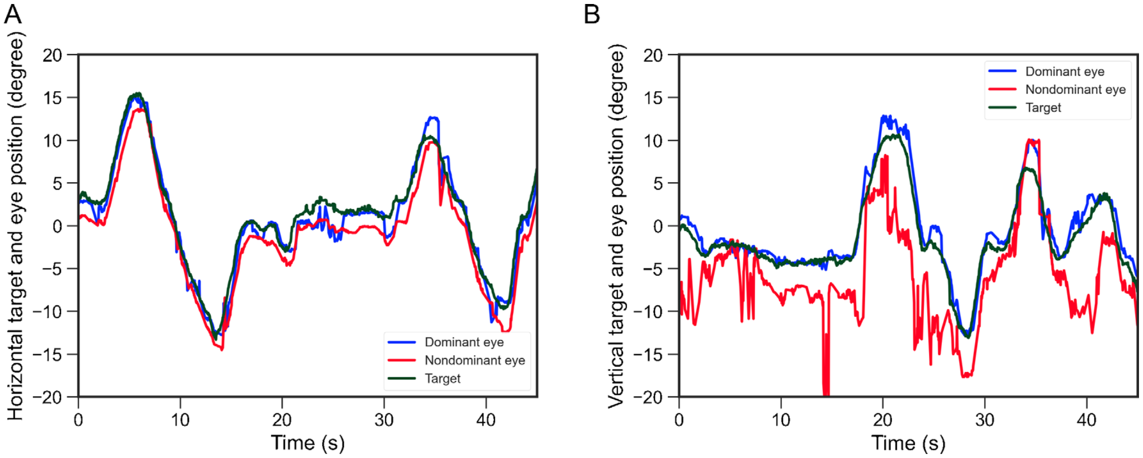

Figure 9. **Horizontal (A) and vertical (B) target locations and eye positions when the examiner moved the target by hand.** The blue, red, and green lines indicate the horizontal dominant and nondominant eye positions and target location during the eye movement test in a patient with postsurgical congenital superior oblique muscle palsy.

Fig. 9A). This finding suggests that the combination of VOG and SSD is suitable for evaluating SPEM in clinic settings. Contrarily, the assessment of the vertical eye movement in the nondominant eye was not accurate due ptosis (see Fig. 9B). This result is a limitation of VOG. The algorithm of VOG that uses the corneal reflex and pupil center failed to detect the pupil due to dark pixels caused by the eyelids and eye lashes.[30] Therefore, we will investigate the SPEMs of patients with strabismus after determining the measurable palpebral fissures in further studies.

SSD analyzed the video recorded during the eye movement test with a mean sampling rate of 32.25 fps, which surpassed the sampling rate of the scene camera. This finding suggests that the combination of VOG and SSD can provide a near real-time analysis of the movements of the hand-held stimulus target and SPEM. The SSD method has the advantage of being able to freely select the targets to be recognized. In this study, 100 epochs of training were completed in 880 seconds. Although the training time for SSD depends on the computer specifications, it can be considered for clinical implementation.

A limitation of our system is the low sampling rate due to the use of a scene camera. If the target moves at a high speed, the target captured by the scene camera becomes blurred, so the accuracy of SSD decreases. Therefore, our system cannot accurately evaluate SEMs. In future research, we plan to update our system to improve the sampling rate of the scene camera so that saccadic eye movement can be analyzed accurately.

Conclusions

SSD achieved high accuracy in recognizing the target that was moved manually, and the target location was significantly and positively correlated with the positions of both eyes as recorded by VOG. Therefore, our findings indicate that the combination of VOG and SSD is suitable for evaluating SPEM in clinical settings.

Acknowledgments

The authors thank Mika Suda for illustrating the mascot character.

Supported by Grands-in-Aid for Early-Career Scientists, Scientific Research (A) and (B), Challenging Exploratory Research, Japan Society for the Promotion of Science (19K20728 [M.H.], 18H04116 [MH], 20K04271 [MH], and 19K21783 [MH], respectively), and Charitable Trust Fund for Ophthalmic Research in Commemoration of Santen Pharmaceutical's Founder (M.H.).

Author Contributions: M.H. conceived the project and designed the experiments. M.H. produced the apparatus. M.H. performed experiments. M.H., T.H., Y.I., E.W., and A.M. analyzed the data. M.H., T.H., and A.M. wrote the manuscript. All authors reviewed the manuscript.

Disclosure: **M. Hirota,** The Patent no. is 2020-207084 (P); **T. Hayashi,** None; **E. Watanabe,** None; **Y. Inoue,** None; **A. Mizota,** None

References

1. Rashbass C. The relationship between saccadic and smooth tracking eye movements. *J Physiol.* 1961;159:326–338.
2. Robinson DA. The mechanics of human smooth pursuit eye movement. *J Physiol.* 1965;180:569–591.
3. Westheimer G, McKee SP. Visual acuity in the presence of retinal-image motion. *J Opt Soc Am.* 1975;65:847–850.
4. Kowler E, van der Steen J, Tamminga EP, Collewijn H. Voluntary selection of the target for smooth eye movement in the presence of superimposed, full-field stationary and moving stimuli. *Vision Res*. 1984;24:1789–1798.
5. Lisberger SG, Morris EJ, Tychsen L. Visual motion processing and sensory-motor integration for smooth pursuit eye movements. *Annu Rev Neurosci*. 1987;10:97–129.
6. Krauzlis RJ. Recasting the smooth pursuit eye movement system. *J Neurophysiol*. 2004;91:591–603.
7. Heinen SJ, Potapchuk E, Watamaniuk SN. A foveal target increases catch-up saccade frequency during smooth pursuit. *J Neurophysiol.* 2016;115:1220–1227.
8. Shanidze N, Ghahghaei S, Verghese P. Accuracy of eye position for saccades and smooth pursuit. *J Vis*. 2016;16:23.
9. Levin S, Luebke A, Zee DS, Hain TC, Robinson DA, Holzman PS. Smooth pursuit eye movements in schizophrenics: quantitative measurements with the search-coil technique. *J Psychiatr Res*. 1988;22:195–206.
10. Imai T, Sekine K, Hattori K, et al. Comparing the accuracy of video-oculography and the scleral search coil system in human eye movement analysis. *Auris Nasus Larynx*. 2005;32:3–9.
11. Ingster-Moati I, Vaivre-Douret L, Quoc EB, Albuisson E, Dufier JL, Golse B. Vertical and horizontal smooth pursuit eye movements in children: a neuro-developmental study. *Eur J Paediatr Neurol*. 2009;13:362–366.
12. Harley RD. Paralytic strabismus in children. Etiologic incidence and management of the third, fourth, and sixth nerve palsies. *Ophthalmology*. 1980;87:24–43.
13. Fukushima J, Tanaka S, Williams JD, Fukushima K. Voluntary control of saccadic and smooth-pursuit eye movements in children with learning disorders. *Brain Dev*. 2005;27:579–588.
14. Lions C, Bui-Quoc E, Wiener-Vacher S, Seassau M, Bucci MP. Smooth pursuit eye movements in children with strabismus and in children with vergence deficits. *PLoS One*. 2013;8:e83972.
15. Metsing I, Ferreira J. The prevalence of poor ocular motilities in a mainstream school compared to two learning-disabled schools in Johannesburg. *African Vision and Eye Health*. 2015;75: a328.
16. Raashid RA, Liu IZ, Blakeman A, Goltz HC, Wong AM. The initiation of smooth pursuit is delayed in anisometropic amblyopia. *Invest Ophthalmol Vis Sci*. 2016;57:1757–1764.
17. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. *arXiv e-prints*; 2015:arXiv:1506.02640.
18. Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector. *arXiv e-prints*; 2015:arXiv:1512.02325.
19. Ren SQ, He KM, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans on Pattern Analysis and Machine Intell*. 2017;39:1137–1149.
20. Cho H, Rybski PE, Bar-Hillel A, Zhang W. Real-time pedestrian detection with deformable part models. *2012 IEEE Intelligent Vehicles Symposium*. 2012;2012:1035–1042.
21. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *arXiv* 2015;arXiv:1512.03385.
22. Szegedy C, Vanhoucke V, Loffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. *arXiv* 2015;arXiv:1512.00567.
23. Kwon KA, Shipley RJ, Edirisinghe M, et al. High-speed camera characterization of voluntary eye blinking kinematics. *J R Soc Interface*. 2013;10(85):20130227.
24. Salvucci D, Goldberg J. Identifying fixations and saccades in eye-tracking protocols. *ETRA '00: Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*. 2000; pp. 71–78.
25. Altman DG, Bland JM. Measurement in medicine - the analysis of method comparison studies. *Statistician*. 1983;32:307–317.
26. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8:135–160.
27. Engel KC, Anderson JH, Soechting JF. Oculomotor tracking in two dimensions. *J Neurophysiol.* 1999;81:1597–1602.
28. Erkelens CJ. Coordination of smooth pursuit and saccades. *Vision Res*. 2006;46:163–170.

29. Rottach KG, Zivotofsky AZ, Das VE, et al. Comparison of horizontal, vertical and diagonal smooth pursuit eye movements in normal human subjects. *Vision Res*. 1996;36:2189–2195.

30. Mulligan JB. Image processing for improved eye-tracking accuracy. *Behav Res Methods Instrum Comput*. 1997;29:54–65.

Supplementary Material

Movie 1. **Representative video of VOG combined with SSD.** The white cross and white square on black squares indicate left and right gazes, respectively, in subject 7. The red bounding box indicates the area that the SSD model recognized as the target.