Motion tracking of iris features to detect small eye movements

Aayush K. Chaudhary
Jeff B. Pelz
Carlson Center for Imaging Science
Rochester Institute of Technology, NY, USA
akc5959@rit.edu, pelz@cis.rit.edu

The inability of current video-based eye trackers to reliably detect very small eye movements has led to confusion about the prevalence or even the existence of monocular microsaccades (small, rapid eye movements that occur in only one eye at a time). As current methods often rely on precisely localizing the pupil and/or corneal reflection on successive frames, current microsaccade-detection algorithms often suffer from signal artifacts and a low signal-to-noise ratio. We describe a new video-based eye tracking methodology which can reliably detect small eye movements over 0.2 degrees (12 arcmins) with very high confidence. Our method tracks the motion of iris features to estimate velocity rather than position, yielding a better record of microsaccades. We provide a more robust, detailed record of miniature eye movements by relying on more stable, higher-order features (such as local features of iris texture) instead of lower-order features (such as pupil center and corneal reflection), which are sensitive to noise and drift.

Keywords: Microsaccades, eye movements, eye tracking methodology, iris features, visual fixations, video-based eye tracking, head motion compensation, iris segmentation.

Introduction

To maintain vision and compensate for the intersaccadic drift between large eye movements, humans continuously make small (miniature) eye movements including tremor, drift, and microsaccades (Cornsweet, 1956; Ditchburn et al., 1959; Nachmias, 1961). These movements occur even when we try to fixate on a point in the world. Low-frequency oscillations of the eyes, unsteadiness of the oculomotor system or attempts to maintain visual perception might result in the production of these movements (Ditchburn & Ginsborg, 1953; Riggs et al., 1953). Among these small eye movements, microsaccades have been an area of interest in the community because of

their vital importance in restoring perception and maintaining vision (Carpenter, 1988; Costela et al., 2013; Martinez-Conde et al., 2006; McCamy et al., 2012; Troncoso et al., 2008), medical diagnosis (Chen et al., 2010; Otero-Millan et al., 2011; Serra et al., 2008) and their ability to indicate attentional shifts (Engbert & Kliegl, 2003; Hafed & Clark, 2002 ; Laubrock et al., 2005).

Microsaccades are the small, jerk-like motions (Martinez-Conde et al., 2004) which are generally found to co-occur in two eyes (Ditchburn & Ginsborg, 1953; Lord, 1951; Møller et al., 2002; Yarbus, 1967, p. 109). While the term 'microsaccade' is sometimes used only for involuntary eye movements that occur during fixation (e.g., Martinez-Conde et al., 2004), we adopt the less restrictive definition of a microsaccade as any rapid eye movement with an amplitude below 0.5 degrees of visual angle. Microsaccades and saccades follow the same *main sequence* (Bahill et al., 1975) and have a common generator (Zuber et al., 1965). Following previous work (e.g., Engbert & Kliegl, 2003; Ko, et al., 2010; Otero-Millan et al., 2008; Troncoso et al., 2008), we regard both voluntary and involuntary saccades below 0.5° as *microsaccades*.

Currently, there is disagreement in the literature about the presence of monocular microsaccades during everyday vision. Gautier et al. (2016) has argued that the practical role of monocular microsaccades may be to aid vergence and make accurate corrections of eye position. However, recent publications (Fang et al., 2018; Nyström et al., 2017) have questioned the existence of such eye movements and raised concerns about previously published results, suggesting that the majority of monocular microsaccades may be measurement artifacts.

One of the limitations of the current generation of video-based eye trackers is their inability to precisely detect microsaccades because the size of those eye movements is similar to the noise level of the trackers which rely on precisely localizing the pupil and/or corneal reflection (CR) on successive frames. Moreover, some of these methods use only pupil information to acquire a higher sampling rate, but in these methods, relative motion between the head and camera are not taken into account (Hermens, 2015), which can increase the false detection of microsaccades. Fang et al. (2018) and Nyström et al. (2017) analyzed many small eye movements that were classified as monocular microsaccades by such algorithms and suggested that most of those events were binocular events or noise. Fang et al. (2018) adjusted the threshold parameter to classify them correctly as binocular microsaccades whereas Nyström et al. (2017) used manual inspection for verification.

New algorithms that take advantage of modern high-resolution cameras and recent advances in computer vision can overcome some of these issues. High-resolution cameras make it possible to extract fine local features such as iris textures. Using populations of such features in place of large, single features such as the pupil allows for a higher degree of confidence in microsaccade detection. Additionally, tracking the motion of those features across consecutive frames allows for the study of motion distributions, simplifying the analysis of small eye movements. We demonstrate that microsaccades less than 0.20 degrees of visual angle can be reliably detected with these techniques and a new velocity tracking algorithm. The method is validated with two types of experimental tasks; reading characters on a test target and watching videos.

Methods

Eye movements are monitored by extracting the motion vectors of multiple iris features. A high frame-rate, high-resolution camera is used to capture a video of an observer's face containing the eyes and surrounding regions. A trained convolutional network (CNN) (LeCun et al., 1998) is used to segment the iris from each frame to extract features only from the region of interest. To ensure high-quality motion signals, Speeded Up Robust Features (SURF) feature descriptors (Bay et al., 2006) are used to extract feature vectors which are then matched in consecutive frames using brute force matching followed by random sample consensus (RANSAC) (Fischler & Bolles, 1981) and homography (Hartley & Zisserman, 2004). Tracking the geometric median of these matched key-points excludes outliers, and the velocity is approximated by scaling by the sampling rate (Pelz & Witzner Hansen, 2017). Microsaccades are then identified by thresholding the velocity estimate. For head motion compensation across each frame, a hybrid cascaded similarity transformation model is introduced using tracking and matching
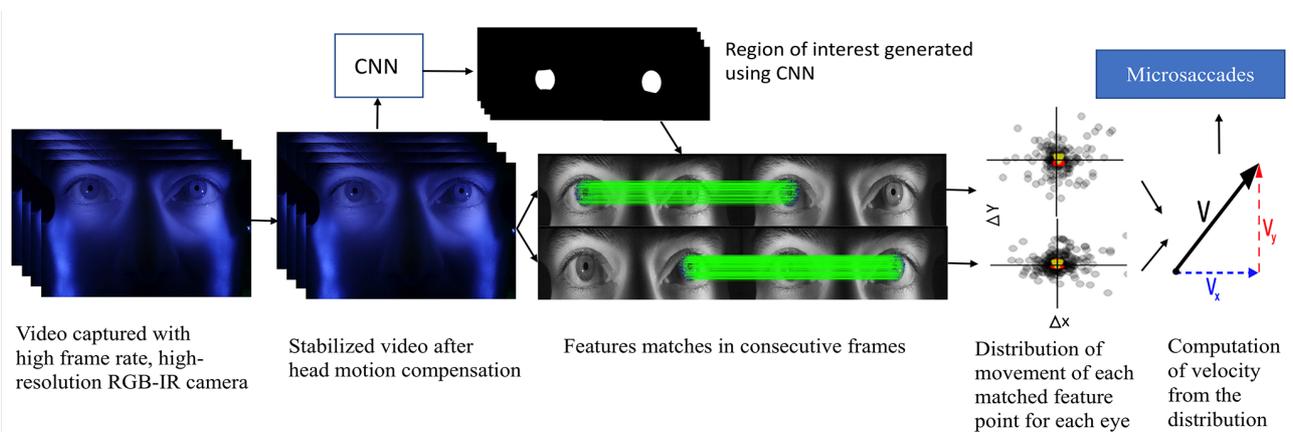

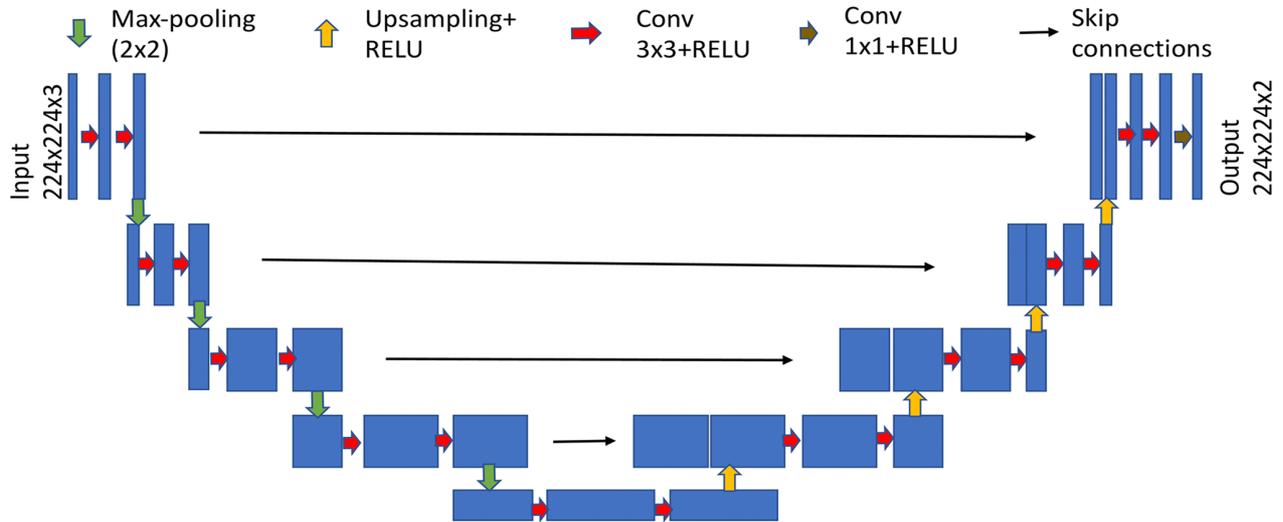Figure 1: Block diagram of the overall system.

Figure 2: Segmentation model based on U-Net architecture after (Ronneberger et al., 2015).

of features of rectangular patches in regions of the observer's cheek. The overall system block diagram is shown in Figure 1.

Iris Segmentation

The region of each frame containing the iris must be isolated from the rest of the frame before velocity estimation can begin. A number of approaches have been used for iris segmentation including ellipse fitting, geodesic active contours (Shah & Ross, 2009), Hough circle fitting, edge detection, integrodifferential operators (Daugman, 1993), graph cuts (Pundlik et al., 2008), and Zernike moments (Tan & Kumar, 2011). All of these methods require tuning for good results and fail to generalize across observers with different iris and skin pigmentation. A trained CNN makes it is possible to generate a more robust solution across observers.

Network Architecture. A U-Net based architecture (Ronneberger et al., 2015) allows for segmentation of the eye region with a moderate training set. The model combines the localization and context information using contracting and expanding paths. The contracting path consists of sequences of two blocks of a 3x3 convolution layer followed by a rectified linear unit (ReLU) as non-linear transformation with batch normalization, then a 2x2 max-pooling operation. The expanding path follows sequences of upsampling with a scale factor of two and then concatenating with its subsequent feature map from a skip-connected layer. This upsampling block is followed by two

blocks of 3x3 convolution layers and later by activation of ReLU with batch normalization. The model achieves good performance and is best suited for biomedical applications with limited data (Ronneberger et al., 2015). Figure 2 shows the segmentation model.

Data labelling. Instead of training on the full-resolution (1920x1080) video, each frame was partitioned into three regions: left and right eye (960x540 each) and the lower part of the face (960x1080). Binary labels of iris and non-iris regions were generated for training the model by manually clicking six – ten points on the border of the iris on each image as shown in Figure 3, and an ellipse was fit to those points with a least square ellipse fitting method (Bookstein, 1979). Since the generated ellipse often overlapped portions of the eyelids, points were also selected along the border with the upper and lower eyelids and used to fit second-degree polynomials to produce the final ground-truth iris regions. The training set contained a total of 406 images from video frames of four observers. The training set was augmented by flipping each image horizontally, producing a total of 812 images. The testing set consisted of 260 images of correlated data (training and test set from the same observers) and 126 images of uncorrelated data.

Training procedure. Training images were re-sized to 224x224. The R, G, and B channels were used as input to the model for the desired two (iris and non-iris) output classes. The Adam Optimizer (Kingma & Ba, 2014) was applied for regularization with a learning rate of 0.0001

Original frame   Left and right eye   Points selection on iris boundary   Remove upper and lower lids   Extracted mask
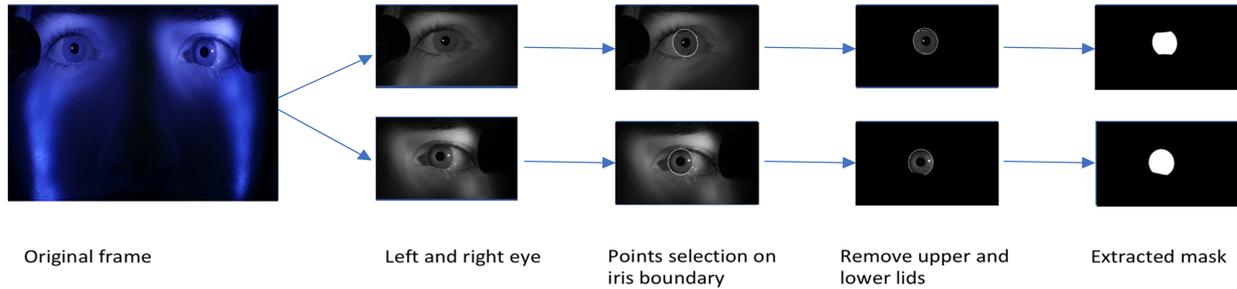
Figure 3: Procedure used for labelling the train and test sets.

with an exponential decay rate for the first moment estimate of 0.55, and an exponential decay rate for the second moment estimate of 0.99. The models were run in Pytorch (Paszke et al., 2017) on an Ubuntu 16.04 LTS platform on a Nvidia Titan 1080 Ti. The training was done with a batch size of eight samples (limited by GPU memory), and the model was run for 40 epochs.

Performance Metrics. Pixel-wise cross entropy of class probabilities was used during training as a loss function. The accuracy of iris detection was measured with the Intersection over Union (IoU) metric. The IoU metric varies from $0 - 1$ and is defined as the ratio of the intersection of the ground-truth and predicted regions to the union of those regions (Everingham et al., 2005). Perfect agreement between ground-truth and prediction yields 1; no overlap yields 0.

If P is the predicted label and G is the ground truth, IoU is computed as shown in Figure 4.

$$IoU = \frac{P \cap G}{P \cup G} =$$

Figure 4: IoU is computed as the ratio of intersection over union. The overlapping regions and non-overlapping regions are indicated by red and gray colors respectively.

Head-motion compensation

Even when an observer's head is stabilized by a chin rest, head movements on the scale of small eye movements still occur, so it is crucial to compensate for this motion when detecting microsaccades. Complex solutions using 3D head models or external hardware trackers could be used, but in this work, we propose an image-based method that compensates with a simple planar image transformation based on regions detected on the face.

The human face is non-planar and deformable. The eyes are not necessarily in the same plane as the nose, cheek or forehead, making planar transformation models like homography (8 degrees of freedom (DOF)), affine (6 DOF) or similarity (4 DOF) transforms imperfect approximations. Here, we make the simplifying assumption that the regions selected in Figure 5 fall approximately in the same plane as that of the two irises being tracked. Relatively planar regions that do not move with normal eye movements were manually selected for each subject.

Figure 5: Indication of regions (light patches) on face tracked along with iris regions for head-motion compensation.

Head movements were compensated by automatically aligning each frame of a video to the initial frame of that video. Computing the homography transformation between frames based on good feature matches across adjacent frames allows a sequence of frames to be aligned. A list of good matches between pairs of adjacent frames was determined with Lowe's ratio test (Lowe, 2001) followed by outlier removal with RANSAC for matched features between consecutive frames. The matched features were found by extracting the local features in the selected regions using SURF and then by using the brute-force matcher where the descriptor of each feature in the nth frame was matched with descriptors of all features in the $(n+1)^{th}$ frame using the L2 distance norm.

Motion compensation can be accomplished for an entire video by aligning each frame in this way to the initial reference frame by cascading the homography transformations as proposed by Dutta et al. (2014). We found, however, that this approach aggregated skew and perspective errors across frames because the points on which computation of the homography matrix was based were not consistent within each region. Because there tended to be more texture in the regions of the lower cheek than near the eyes even though they were in approximately the same plane, more of the features used for tracking were from the cheek regions. Constraining the feature path to a rectangle as in Grundmann et al. (2011) decreased errors but did not provide an acceptable solution.

Instead, we selected features in rectangular regions as seen in Figure 6, then computed good matches between consecutive frames for each of the regions. Computation of the geometric median of those matches provided an estimate of the best value of the distribution of the rectangular patch in the source and destination images. Using one point from each of the four regions we computed the transformation between image pairs. The homography transformation is the least-constrained plane-to-plane mapping, with eight degrees of freedom (DOF) which can be described as allowing translation (2 DOF), scaling (2 DOF), rotation (2 DOF), and keystoning (magnification varying by position; 2 DOF). Other transformations are more constrained. The affine transform, for example, allows only six degrees of freedom which can be described as translation (2 DOF), scaling (2 DOF), rotation (1 DOF), and shearing (translation varying by position; 1 DOF). The similarity transformation has only four degrees of freedom, allowing only translation (2 DOF), uniform scaling (1 DOF), and rotation (1 DOF). As shown by Grundmann et al. (2011), the extra degrees of freedom of the homography and affine transformations can result in larger errors

Figure 6: Regions (light patches) on four observers faces used for head-motion compensation.

than the more constrained similarity transformation, but we found that skew errors still grew over time even with the similarity transform.

Dutta et al. (2014) identified *breaks* in a video sequence when the video starts panning or scale changes become substantial. Information from prior frames is not considered after a *break* and a new reference frame is introduced. Grundmann et al. (2011) proposed a different approach in which all the keyframes are matched, and transformation between the keyframes is initially computed. Then, a similarity transformation is found from the starting keyframe to each subsequent frame until the next keyframe. A small jitter is observed when moving from the final intermediate frame to the next aligned keyframe, so there is a tradeoff in the frequency of re-aligning with a new keyframe; more frequent realignment minimizes the size of the jitter but increases its frequency.

We elected to realign to a new keyframe once every 480 frames (5 s). Less frequent realignment increased the aggregated skew and perspective errors, and more frequent realignment increased the frequency of the jitter. The overall model is shown in Figure 7.
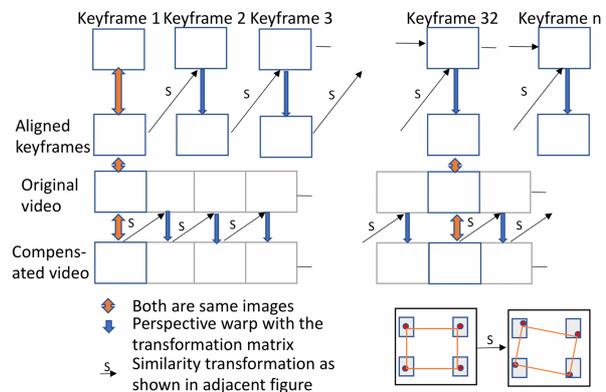

Figure 7: Model used for head motion compensation.

Velocity approximation

After the segmentation of irises in the stabilized video, the next crucial step is to extract the iris features. These features were extracted in a Contrast Limited Adaptive Histogram equalized grayscale representation of the image (Pizer et al., 1987) as this resulted in a larger number of good matches across different observers than did any single-color channel. From the extracted features across consecutive frames, good matches were found using brute-force matching followed by Lowe's ratio distance test and
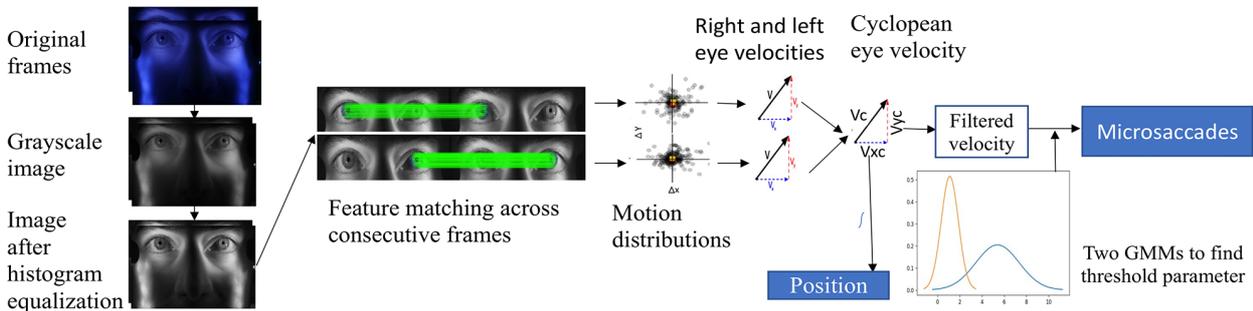
Figure 8: Block diagram of iris feature extraction to microsaccades detection.

outlier removal by RANSAC. The motion of the iris between adjacent video frames was represented as a motion vector.

These motion vectors can be used to compute horizontal, vertical and torsional eye movements. A number of researchers have used iris features to measure torsional eye movements (Groen et al., 1996; Lee et al., 2007; Ong & Haslwanter, 2010). Ong & Haslwanter (2010) transformed the image of the iris into polar coordinates about the pupil center so that torsion could be tracked as translation along the angular axis.

Because we are concerned with microsaccades in this work, we ignored torsional movements and computed the geometric median of the shift of all matched keypoint pairs between frames. The relative velocity of each eye was calculated independently as the product of the scaled motion vector and the sampling rate, and integrating those values gave a relative position signal for each eye. Calibration was performed by simple linear regression of the relative position signal and the known gaze position during calibration. Position cannot be estimated during blinks, so the data reported here include only trials completed without blinks during the calibration routine.

The raw signal was filtered using 1D total variation denoising (TVD) (Condat, 2013) which minimizes the variation as a least squares problem. Instead of computing an estimate on a local neighborhood, TVD uses the entire record to estimate a global optimum (Pekkanen & Lappi, 2017) to minimize undesired spurious noise while preserving the saccadic 'edges.' TVD is preferable to smoothing algorithms because our goal was to model the sensor noise rather than eye position noise using the overall signal. The key parameter in the TVD filter is a regularization parameter which was set to 0.1 for all observers. After denoising,

cyclopean gaze velocity was computed from the right and left eye velocities.

Microsaccade detection

A number of methods have been proposed for microsaccade detection (Bellet et al., 2018; Engbert & Kliegl, 2003; Herrington et al., 2009; Otero-Millan et al., 2014). Almost all of the algorithms have a tunable parameter that affects algorithm performance. We implemented an adaptive version of the Velocity-Threshold Identification algorithm (I-VT) (Salvucci & Goldberg, 2000). Here, microsaccades were detected by thresholding the generated absolute cyclopean velocity signal. Determining the velocity thresholding parameter is an essential step in categorizing events, as a high threshold would increase the miss rate, and a low threshold would increase false alarms. A fixed threshold did not work well for all observers because of individual variations, noise in the system due to prevailing signal artifacts and small, uncompensated head movements.

To automatically adapt to individual observers, we implemented an adaptive-threshold method based on a Gaussian mixture model (GMM) with two velocity distributions representing noise and microsaccades. To eliminate larger saccades, only absolute velocities below 20 deg/sec were passed to the model. A thresholding parameter based on estimated mean and standard deviation using 99.7% of each distribution was used. Based on observations of the noise floor, a lower bound of 3.84 deg/sec was selected to minimize false alarms. Figure 8 shows the block diagram from the iris feature extraction to microsaccade detection.

Experimental Setup

Binocular eye movements were recorded using a Panasonic Lumix DMC-GH4 mirrorless digital camera modified by removing the IR-rejection filter. Video recordings of the eyes and the region surrounding the eyes were recorded at a frame rate of 96 frames per second (fps). Head movements were restricted with a UHCOTech HeadSpot chin and forehead rest. Filtered tungsten and LED infrared light sources were used to illuminate the eyes and the region surrounding the eyes as shown in Figure 9. Subjects were asked to sit comfortably to minimize head movements. Trials were repeated until calibration was completed without blinks.

Apparatus. The Lumix DMC-GH4 was set to ISO 400, and F/8. The camera was placed at a distance of 50 cm from the observer, and the center of the lens was positioned 4.5 cm below the eyes at an angle of approximately 4 degrees with the horizontal axis (the frame was not centered on the eyes). A focal length of approximately 70 mm was used to capture an appropriate region including the eyes and surrounding areas.
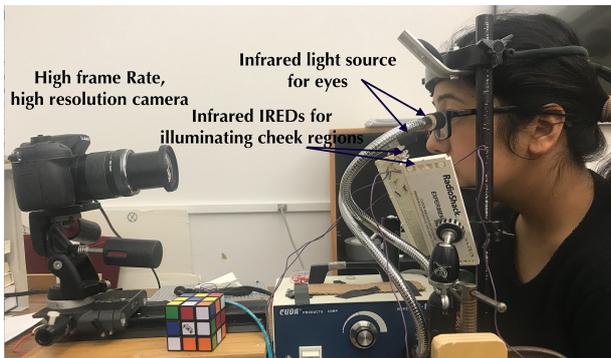

Figure 9: Experimental setup with calibration targets and the scene behind the camera.

Light Source. The experiments were conducted in a lab with indirect fluorescent illumination. A tungsten-halogen source with a bifurcated fiber-optic light guide was filtered with a near-infrared high-pass (750nm) filter. As seen in Figure 9, the light guides were placed at an angle ~25-30 degrees above the horizontal, separated by 88mm. The sources provided an irradiance of 0.020 – 0.037 W/cm^2 at the iris. Each side of the observers' cheek regions was illuminated with twelve 940 nm infrared emitting diodes (IREDs) (Vishay VSLY5940) adjusted to match the exposure of the iris. The IREDs were arranged in parallel lines, separated by 7 mm, and placed to the side of the observer so that the topmost IREDs were approximately 5 cm below the eyes.

Subjects

Eye movements of seven participants (5 males, 2 females) with normal or corrected-to-normal vision were recorded. Participants with light and dark irises and with and without prescription glasses were selected. Subjects were undergraduate and graduate students with a mean age of 25 years (σ=3). The experiment was conducted with the approval of the Institutional Review Board and with the informed consent of all participants.

Tasks

Each observer performed three tasks, each preceded by a task-specific 9-point calibration routine.

Snellen microsaccade task

Microsaccades can be evoked when observers read small, isolated characters, as in the Snellen eye chart (Shelchkova et al., 2018). Calibration points and a 'pocket' Snellen chart (7″, x 4″, designed for use at 14″) were placed at a distance of 150 cm from the observer 3 cm above the center of the subject's eyes. The field of view of the calibration target was 8.7° x 6.9°. The subjects were asked to fixate each point in the calibration target, then look at each character on a line of the Snellen chart at a distance of 150 cm. Each character subtended a vertical angle of approximately 5 arcminutes (equivalent to a '20/20' character at that distance). The subtended horizontal angle between eight characters in the chart was approximately 0.15 degrees; two others were 0.28 and 0.32 degrees. The small eye movements evoked as observers looked at the characters on the Snellen eye chart which were used to verify the detection of microsaccades using our algorithm.

Video Stimuli

To examine the detection of microsaccades while viewing a moving stimulus, observers viewed two short videos from (Kurzhals et al., 2014) on a 12-inch Apple MacBook (MF855LL/A). In the first video, the observers were instructed to track the motion of a car turning in a tight *figure-8* pattern. In the second video, they were instructed to track the cup containing a marble in a 'shell

game.' Both videos induced smooth pursuit eye movements. The laptop was placed on a table so that the display center was 100 cm from the observer and 10 cm above the subject's eyes. The display field of view was 13.5° X 7.4°.

Results

Segmentation Results

Figure 10 shows the loss function for the training and the testing datasets. The model starts to reach its asymptote at 35 epochs for both sets. For our model, the average IoU values were 0.898 for the training set, 0.891 for the correlated test data, and 0.866 for the uncorrelated test data.
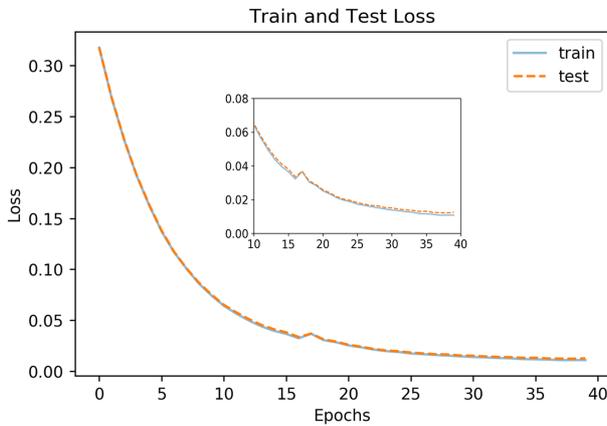

Figure 10: Loss curve for segmentation model.

Figure 11 shows images with iris regions predicted by the model and ground truth indicated in yellow and green colors respectively. For proper visualization, only the border predicted by the model is shown. The large white regions in the upper-right image are specular reflections in the observer's eyeglasses. A spike in the loss function at epochs 16-17 evident in Figure 10 is because of inconsistency in some of the labeled data such as when the eyelids cover the iris as seen in the lower left image in Figure 11.

Video Stabilization

We evaluated the video stabilization for two cases; a rigid Styrofoam model head and for real human faces. For the rigid head model, we tracked the four points labeled A, B, C and D in Figure 12 (left). The rigid head was rotated slightly about the base, resulting in horizontal, vertical and diagonal movements of approximately 2 mm at a frequency of approximately 1 Hz, and the video was stabilized using our compensation model. The original motion for point B is shown in solid red in Figure 13 and the motion-compensated output is shown in dashed blue.
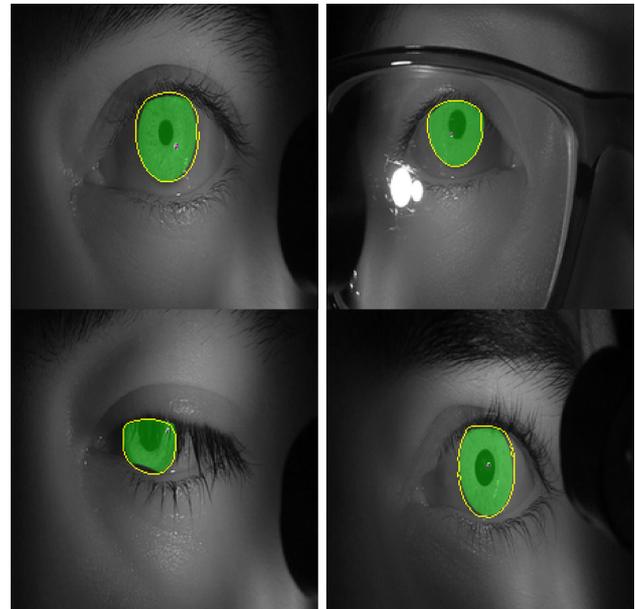

Figure 11: Images of test set with labels predicted by the model in yellow color (edges) and the labelled ground truth in green color (filled).
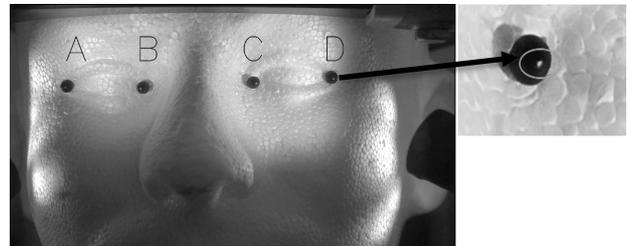

Figure 12: Face model used for head stabilization verification with markers A, B, C and D indicated in the figure (Left). (Right) The circle indicates the bright spot used for our method verification.

Table 1 shows the mean squared change (MSE) from the initial starting reference frame and standard deviation (STD) of points before and after stabilization in pixels. The overall mean per class (MPC) across all points shows a dramatic improvement in performance. The test points were tracked by locating the brightest spot (a specular reflection) on the black marker as indicated in Figure 12 (right). A variation of one pixel is expected because the maximum location algorithm in OpenCV (Bradski & Kaehler, 2000) returns the horizontal and vertical pixel location as integer values.

We also examined the performance of the video stabilization algorithm for a seated observer in a chinrest looking at nine calibration points for about a second each. Head motion was tracked by attaching two small stickers to the observer's face and tracking each sticker in the video by computing the mean of the central votes of matched features by using consensus-based matching and tracking (Nebehay & Pflugfelder, 2014). Before and after compensation results are shown in Table 2. Unlike the rigid head model, where intentional movements were introduced, we expected only minimal head movements in the human face condition. The result shows an improvement in most of the

cases, but in one case (S2), the standard deviation increased along the x-axis by approximately 20% while decreasing along the y-axis by 32%.

Table 1: Motion before (without compensation) and after (with compensation) results for each point and overall mean per class (MPC).

Points	Before		After	
	MSE	STD	MSE	STD
A x	5.28	7.19	0.48	0.65
A y	2.62	3.58	0.44	0.51
B x	5.38	7.37	0.61	0.79
B y	1.31	1.8	0.21	0.44
C x	5.53	7.57	0.52	0.76
C y	0.77	1.16	0.51	0.56
D x	5.56	7.64	0.42	0.66
D y	2.26	2.92	0.21	0.44
MPC	**3.59**	**4.9**	**0.42**	**0.6**

Table 2: Standard deviation (in x and y pixel position) of movements of two points in either side of face in the original video and after compensation.

Subject	Sticker placed on	Original		After	
		STDx	STDy	STDx	STDy
1	Left	2.2	1.22	1.05	1.11
	Right	1.45	1.37	1.12	1.03
2	Left	0.8	3.15	0.88	2.23
	Right	1.02	2.49	1.24	1.71
3	Left	0.98	1.67	1.26	1.51
	Right	1.46	2.29	0.57	1.08
4	Left	2.78	5.08	2.9	2.21
	Right	3.11	4.53	1.39	2.24

Snellen microsaccade task

In the first task, observers were instructed to read a line on the pocket Snellen eye chart whose characters subtended an angle of approximately 5 arcminutes (equivalent
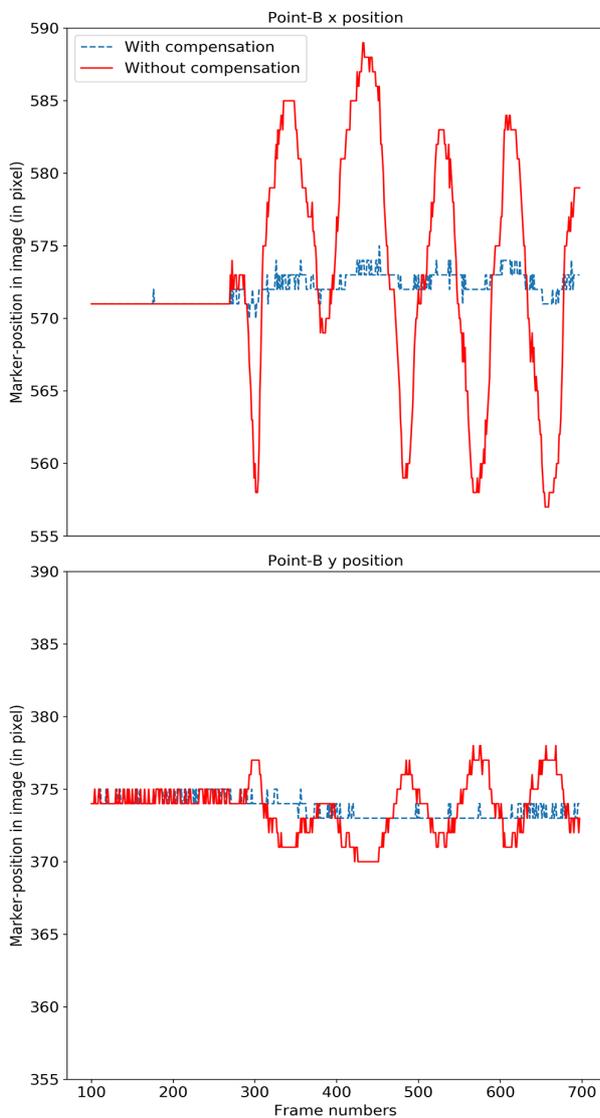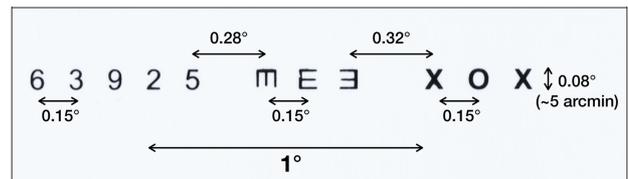

Figure 13: The motion of the original face model (red) and after compensation (blue) for point B is shown in the x-direction (top) and y-direction (bottom).

Figure 14: Target line from pocket Snellen chart. Characters in three groups subtended a vertical angle of 5 arc minutes.

to a '20/20' character at that distance). Because of the size and distance of the target, the eye movements necessary to foveate each target constitute microsaccades (Shelchkova et al., 2018). Figure 14 shows the target and the angular subtense of the microsaccades required to move between the characters in each of the three groups (0.15°) and between the two groups (0.28° and 0.32°). Figure 15 shows the events detected for one of the subjects. The blue line indicates the cyclopean eye absolute velocity; the green dashed lines indicates the microsaccades or drifts detected which is a superset of the fixation targets. These microsaccades were detected with an absolute velocity threshold of 3.84 deg/sec. Note the clear separation between the signal and noise. Red lines indicate saccades with an absolute velocity greater than 50 deg/sec. The saccadic movement that appears every 5000ms is an artifact of the jitter caused when a new keyframe is taken as a reference during head stabilization.

Figure 16 plots the absolute velocity of the cyclopean gaze (with its amplitude), X-t, Y-t and YX graphs for both eyes when microsaccades were detected. The amplitude of each microsaccade was calculated based on the gaze position between the local minima of the cyclopean velocity before and after the microsaccades. The two local minima are indicated in the absolute velocity plot by the vertical light blue lines. There is a close relationship between the two eyes for most of the cases. The center column shows a case where the eyes were moving in opposite directions.

Results for all seven subjects are shown in Table 3. The table shows the number of microsaccades per observer: (A) found when the eye video was inspected visually; (B) detected by the algorithm; (C) detected by the algorithm

when separated by an interval equal to at least half of the average duration available for each character; and (D) detected by the algorithm but not present ('false alarms').

The events for 0.15 degree microsaccades in Table 3 (C) should be (4, 2, 2) as the instructed targets are set in that order as seen in Figure 14. Initially, two events of 0.28-degree and 0.32-degree microsaccades are determined. Then, the number of events before 0.28-degree microsaccades, between 0.28 degree and 0.32 degree and after 0.32 degree are detected with a consideration that only one unique microsaccade must be a voluntary movement made for the target in the interval. All the other movements in the interval are neglected as our interest lies in detecting movements for the instructed target rather than observing the number of overall detected microsaccades. Note that the method identified all small movements ≥ 0.2 degrees.

Table 3: Number of microsaccades detected manually by inspecting video and using the algorithm for all subjects. The total number of instructed targets is 8 targets of 0.15 degree and 2 targets over 0.2 degrees. The table also shows number of false alarms observed.

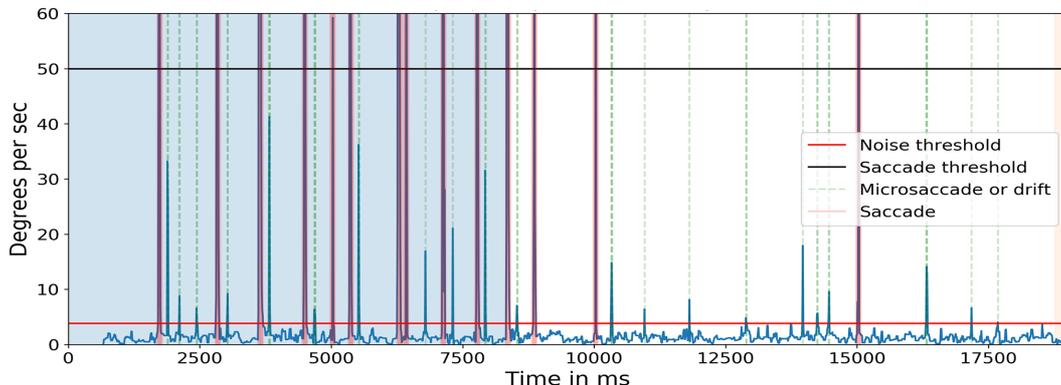|	Subjects		1	2	3	4	5	6	7
0.15 degrees (out of 8 instructed targets)		(A) Microsaccades observed by inspecting video	8	8	3	14	19	15	10
	Algorithm	(B) Detected	6	8	3	16	18	12	10
		(C) Finite interval gap (out of (4,2,2) events)	(2,1,2)	(4,1,2)	(2,1,0)	(4,0,5)	(5,1,6)	(3,6,2)	(3,3,1)
(D) Number of false alarms			0	0	0	2	1	2	0

Figure 15: Absolute velocity (deg/sec) of the cyclopean gaze. Light green shows the small eye movements detected between 3.84 deg/sec to 50 deg/sec. There should be at least eight 0.15-degrees, one 0.28-degrees, and one 0.32-degrees, a total of 10 movements after 8500ms (in the non-highlighted region) as targets in one line of the Snellen chart. The initial 8500ms are the calibration phase.
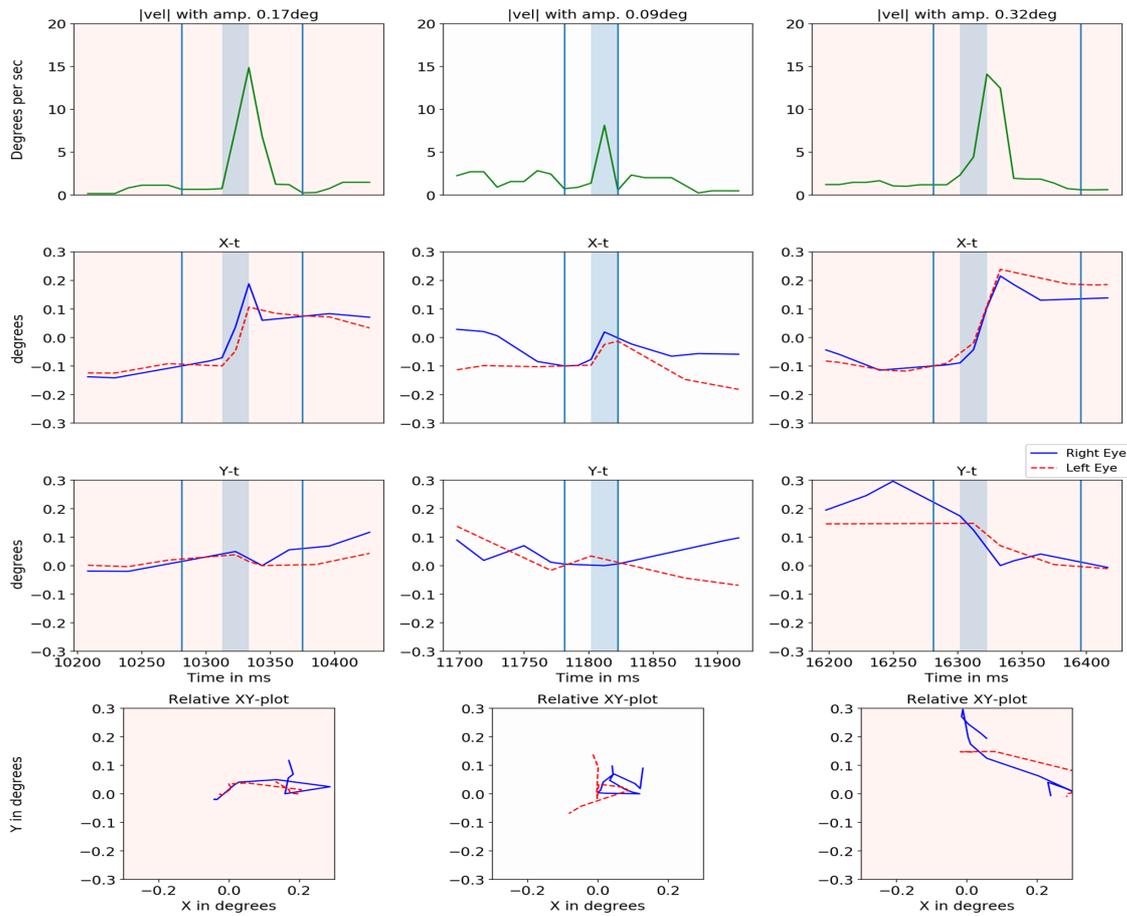
Figure 16: Cyclopean velocity plot, relative X-t, Y-t and XY plot for right eye (blue) and left eye (dashed red) is shown for three events in different columns.

Identifying small movements with a magnitude of 0.15 degree is possible for some events when a lower threshold is selected, but this increased the false alarm rate. Subjects 5 and 6 were two cases of noisy data. For Subject 6, using an initial regularization parameter of 0.1 for filtering resulted in eight false alarms and a miss rate of 11.1% (with respect to the total number of microsaccades observed by inspecting video). It was seen that signal was noisy even after head motion compensation thus changing filtering regularization parameter to 0.20 and 0.15 for the right and left eye respectively there was an improvement in the result with only two false alarms. However, higher values of filtering regularization parameter increased the miss rate to 33.3%.

The importance of compensating for small head movements can be seen in Figure 17. Without compensation (top panel), a threshold of 12.0 deg/sec was necessary to exclude noise. After compensation (bottom panel), a
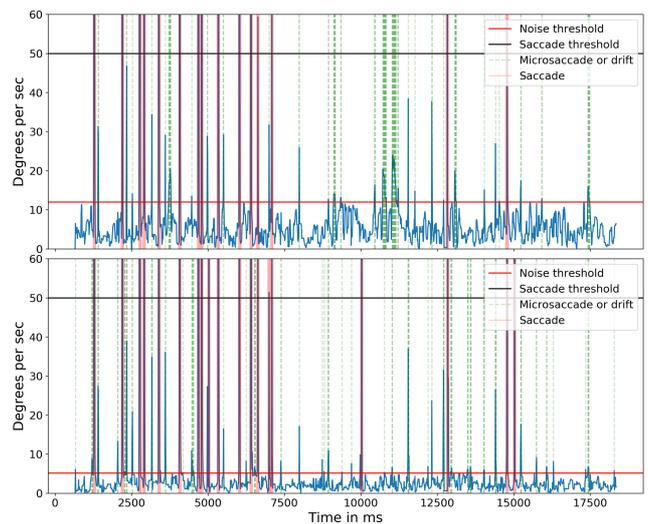

Figure 17: Absolute velocity (deg/sec) of the cyclopean gaze plotted for subject 5 for both without head compensation (top) and after head compensation (bottom). Light green shows the small eye movements detected for their respective threshold parameter.
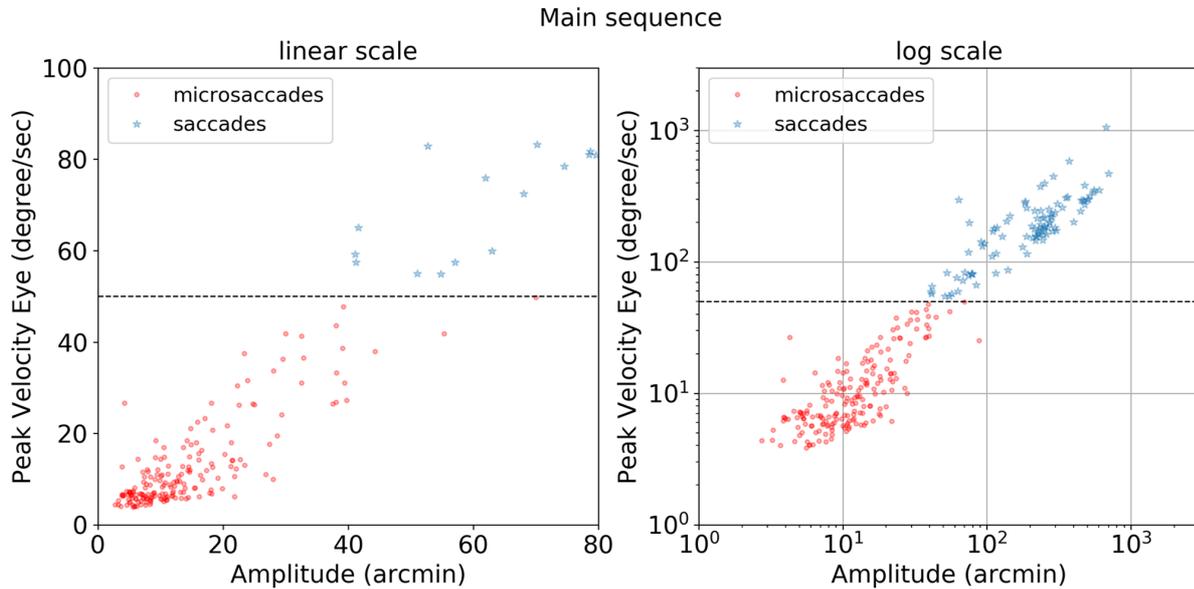
Figure 18: Main sequence plot of peak velocity (degree/sec) vs amplitude (arcmin) in linear and log scales for microsaccades (red circles) and saccades (blue stars). The horizontal dotted line at 50 deg/sec is a fixed threshold used to differentiate saccades and microsaccades.

threshold of 5.1 deg/sec allowed microsaccades to be detected more accurately. The one false alarm detected for the compensated data was also detected in the uncompensated data indicating a false alarm resulting from large head motion. Without compensating for head motion, a total of nine microsaccades were missed. Figure 18 shows the 'main sequence' (peak velocity vs. amplitude) for microsaccades and saccades detected by our method. Both saccades and microsaccades fall on the main sequence with a slope of 47 s^{-1}. Microsaccades are defined in this context as all movements with a peak velocity <50 deg/sec.

Video Stimuli

In the first video stimulus task, the observers were instructed to fixate on a moving car, leading to smooth pursuit, fixations, saccades, and microsaccades. The box plot with the number of microsaccades detected per second and the variation of the rate of microsaccades among different subjects is plotted for all subjects in Figure 19 (top row). The median number of microsaccades in the video is approximately one per second for most of the trial.

In the second video task, the observers viewed a 'shell game' with three cups and one marble and were instructed to follow the cup that contained the marble. The visual task induced smooth pursuit, saccades, fixations, and microsac-

cades. The box plot with the number of microsaccades detected per second and the variation of the rate of microsaccades among different subjects is plotted for all subjects in Figure 19 (bottom row). Figure 20 compares the average amplitude of the microsaccades over the entire video for the two video tasks.

Discussion

Current video-based eye-tracking methods can lead to ambiguities and potential errors when tracking very small eye movements. Our work represents an advance in analyzing microsaccades by tracking iris textures using a high frame rate, high-resolution camera. We segmented the iris from the video frame with a trained CNN and measured the model accuracy using an IoU metric that rewards correct matches and penalizes false matches. The IoU metric was appropriate because we calculated frame-to-frame velocity based on the geometric median of the population of iris feature matches rather than relying on extracting the precise iris boundary. Our CNN model was able to generalize for a varying set of data, though some labeling errors (e.g., eyelids covering the iris) resulted in inconsistencies. The CNN was trained with labels from a single human labeler and might be improved further if labels with multiple-labeler agreement were used for training.
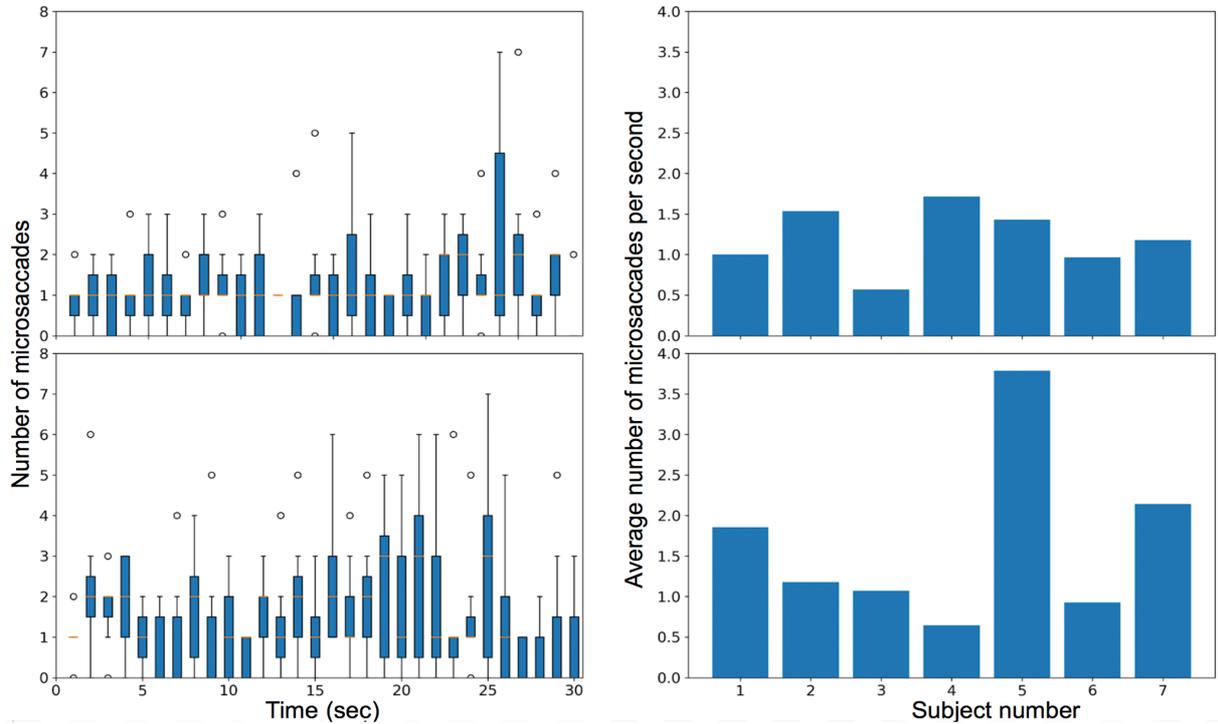
Figure 19: Plot for the rate of microsaccades during trial (left) and the average microsaccades rate for each subject in their overall video (right) when subjects were watching a video of a moving car (top) and watching a video of a shell game (bottom).

Head motion impacts the number of detectable microsaccades, so we introduced a head compensation model based on simple planar transformations. The model achieved excellent performance with a rigid head model, reducing the overall MSE from 3.59 pixels from the initial reference frame to just 0.42 pixels with standard deviation decreasing to $1/8^{th}$ of the original value. The model also improved the detection of microsaccades for human subjects as the number of false alarms decreased significantly after compensating for head motion. However, some false alarms remained in cases where a large jitter was observed because of transformation issues (warping). For some subjects, head movements were comparatively large causing more noise in the signal resulting in higher thresholding value obtained from GMMs for microsaccade detection. The current implementation can miss some microsaccades if they occur while updating to a new reference keyframe. When a new reference keyframe was updated there was usually a significant apparent movement, and the motion was misinterpreted as a saccade. Since our major concern lies in detection of microsaccades when a person fixates the instructed targets at a regular interval (approx. 1 sec) and we did not wish to include post-saccadic oscillations
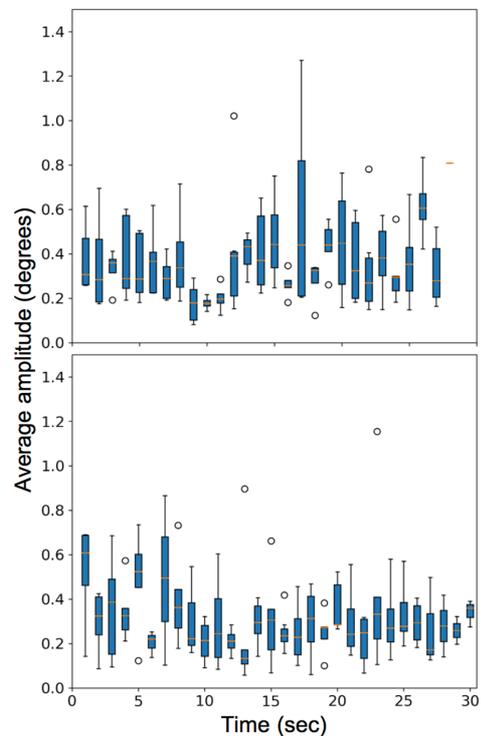

Figure 20: Plot for the average amplitude of microsaccades (degrees) made by various observers for two video stimuli task. (Top) is for watching a video of a moving car and (Bottom) is for watching a shell game.

or movements compensating for overshoots or undershoots as microsaccade events, we required a minimum period of 52ms (5 frames) between detected velocity peaks. In our tests, one microsaccade was missed during that time interval as the microsaccade timing was aligned to the timestamp of keyframe update.

We have demonstrated the robust detection of microsaccades by tracking the velocity signal generated from iris textures, extracting microsaccades with GMM-based adaptive velocity threshold parameter with a mean of 5.48 deg/sec (σ=1.16). In four of 21 trials in three tasks, the threshold value used was the lower bound of 3.84 deg/sec. Note that for head-fixed data (Fang et al., 2018) referred to a threshold of 10 deg/sec as a relatively strict criterion and used the threshold value of 5 deg/sec to validate some monocular events. Our algorithm makes it possible to use this range of parameters (>3.84 deg/sec) consistently without significantly impacting the false alarm rate. 100% of the voluntarily generated small movements over 0.2 degree were detected using our algorithm, and even microsaccades with an amplitude of 0.15 degrees were detected over 73% of the time (41 out of 56 events). Further, microsaccades as small as 0.09 degree were recognized in low-noise backgrounds.

Manually inspecting the video showed that the number of microsaccades was less than the number of instructed targets for the smallest (0.15 degree) microsaccades. It is possible that not all instructed eye movements were actually made; note that an entire grouping of characters fell within the fovea. For three out of seven subjects, all of the 0.15° microsaccades that were identified via visual inspection of the video record were also detected by the algorithm. For one subject, two microsaccades that were not detected by the algorithm were observed to be smaller than the remaining microsaccades for the same observer, perhaps because the preceding microsaccades 'overshot' the previous character. It is likely that individual fixations fall on different positions within each character, and some of the characters subtended a horizontal angle of only 0.04 degrees, so it is possible that the individual microsaccades were less than 0.15-degrees if the eye movements were not between character centers. Thus, the actual magnitude of the microsaccades were not exactly the expected value of 0.15, 0.28 and 0.32 degrees calculated from the original target angles. Note too that our algorithm is best suited for event detection rather than providing an exact estimate of position.

Figure 16 (second column) showed one of the sample events where the two eyes moved in opposite directions. This could be the result of true uncorrelated motion (such as vergence movements) or errors in the head compensation model. Vergence was detected by our method especially before and after the blinks as the eyes converged before a blink and diverged after.

The results for our study for the number of microsaccades per second for the two video stimuli tasks are in agreement with Martinez-Conde et al. (2004) and the number and amplitude range are also consistent with the natural scene, picture puzzle and Where's Waldo head-fixed experiments conducted by Otero-Millan et al. (2008).

Analysis of the video *figure-8* car stimulus suggest that the person is making more microsaccades in the later part of the video. This may be because the car was closer to the camera at that point and the number of catch-up saccades less than 50 deg/sec is expected during smooth pursuit (Laurutis et al., 2010). We also observe that a high variance of movement is made at the 16th and 24th sec where the car makes a rapid movement and covers approximately 1/6th of the visual field while during the rest of the video the car mostly covers approximately 1/15th of the visual field. During the 13th sec, a long pursuit is expected, and we observed a variation in the number of microsaccades detected since some of the subjects made a large amplitude microsaccade, and few made saccades (>50 deg/sec) in that time interval. In the shell game, it was observed that the number of microsaccades increased from the 11th to the 25th-sec interval even when the glass was at rest perhaps because the person was tracking other objects (hand) in the scene that were still moving.

In the video tasks, the threshold was set for each trial by a GMM. As the smooth pursuit velocity induced by the moving targets was less than the microsaccade velocity threshold, the static GMM was sufficient. In the future, we will implement a continually variable adaptive-threshold in which the GMM will be based on a sliding window so that microsaccades superimposed over higher velocity smooth pursuit can be reliably detected without increasing false alarms during the pursuit.

While the Snellen chart offers a convenient target to induce microsaccades, it presents several challenges. At this scale, several targets fall well within the fovea, so an observer doesn't need to make saccades to read each char-

acter. In addition, significant undershoots and/or overshoots could occur that would make it difficult to identify individual eye movements, without affecting an observer's ability to read each character. We used the finite interval in an attempt to reduce errors for localizing individual eye movements, but it is still possible that we were not able to accurately parse each microsaccade perfectly.

This experiment could be improved by using a high-resolution digital display in place of the printed Snellen target, and display only one character at a time. This would simplify the identification of each microsaccade by synchronizing presentation and detection times. A higher frame-rate camera could also improve performance. The temporal resolution of the camera impacts the estimate of the amplitude, as it is obtained by integration of velocity over time. With more discrete velocity samples during brief microsaccades we can estimate the amplitude of microsaccades more accurately. Additionally, the signal-to-noise ratio both in the head compensation model and the iris feature tracking method can be improved by using a higher frame rate camera with all-intraframe (ALL-I) encoding. We are also exploring hybrid algorithms that merge velocity and position signals to gain the precision benefit of the iris motion tracking while maintaining the traditional positional information.

In summary, the main contributions of this paper are the use of trained CNNs to provide a more robust solution for iris segmentation across observers with different iris and skin pigmentation; an image-based model for head motion compensation using planar transformations which can be applied in various applications like video stabilization; extracting high quality iris images rich in textures; and an algorithm to reliably detect small eye movements over 0.2 degrees with very high confidence by extracting motion signals with a high signal-to-noise ratio by computing motion distributions rather than relying on precise pupil boundary localization as in pupil and Pupil-CR systems. This method can identify even smaller movements if the velocity of the movement is higher than the threshold parameter (>3.84 deg/sec).

Ethics and Conflict of Interest

The author(s) declare(s) that the contents of the article are in agreement with the ethics described in http://biblio.unibe.ch/portale/elibrary/BOP/jemr/ethics.html.

JBP is co-inventor of the base method used to extract velocity signals from iris texture features (Pelz & Witzner Hansen, 2017).

References

Bahill, A. T., Clark, M. R., & Stark, L. (1975). The main sequence, a tool for studying human eye movements. *Mathematical Biosciences*, *24*(3-4), 191-204.

Bay, H., Tuytelaars, T., & Van Gool, L. (2006, May). Surf: Speeded up robust features. In *European conference on computer vision* (pp. 404-417). Springer, Berlin, Heidelberg.

Bellet, M. E., Bellet, J., Nienborg, H., Hafed, Z. M., & Berens, P. (2018). Human-level saccade detection performance using deep neural networks. *Journal of neurophysiology.*

Bookstein, F. L. (1979). Fitting conic sections to scattered data. *Computer Graphics and Image Processing*, 9(1), 56-71.

Bradski, G., & Kaehler, A. (2000). OpenCV. *Dr. Dobb's journal of software tools*, 3.

Carpenter, R. H. (1988). Movements of the Eyes, *2nd Rev. Pion Limited.*

Chen, A. L., Riley, D. E., King, S. A., Joshi, A. C., Serra, A., Liao, K., ... & Leigh, R. J. (2010). The disturbance of gaze in progressive supranuclear palsy: implications for pathogenesis. *Frontiers in neurology*, 1, 147.

Condat, L. (2013). A Direct Algorithm for 1-D Total Variation Denoising. *IEEE Signal Process. Lett.*, *20*(11), 1054-1057.

Cornsweet, T. N. (1956). Determination of the stimuli for involuntary drifts and saccadic eye movements. *JOSA*, 46(11), 987-993.

Costela, F. M., McCamy, M. B., Macknik, S. L., Otero-Millan, J., & Martinez-Conde, S. (2013). Microsaccades restore the visibility of minute foveal targets. *PeerJ,* 1, e119.

Daugman, J. G. (1993). High confidence visual recognition of persons by a test of statistical independence. *IEEE transactions on pattern analysis and machine intelligence*, 15(11), 1148-1161.

Ditchburn, R. W., Fender, D. H., & Mayne, S. (1959). Vision with controlled movements of the retinal image. *The Journal of physiology*, 145(1), 98-107.

Ditchburn, R. W., & Ginsborg, B. L. (1953). Involuntary eye movements during fixation. *The Journal of physiology*, 119(1), 1-17.

Dutta, R., Draper, B., & Beveridge, J. R. (2014, March). Video alignment to a common reference. In *Applications of Computer Vision (WACV)*, 2014 *IEEE Winter Conference* on (pp. 808-815). IEEE.

Engbert, R., & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision research*, 43(9), 1035-1045.

Everingham, M., Van Gool, L., Williams, C., Winn, J., & Zisserman, A. (2005). Pascal visual object classes challenge results. *Available from www. pascal-network. org*, 1(6), 7.

Fang, Y., Gill, C., Poletti, M., & Rucci, M. (2018). Monocular microsaccades: Do they really occur? *Journal of vision*, 18(3), 18-18.

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381-395.

Gautier, J., Bedell, H. E., Siderov, J., & Waugh, S. J. (2016). Monocular microsaccades are visual-task related. *Journal of vision*, 16(3), 37-37.

Groen, E., Bos, J. E., Nacken, P. F., & De Graaf, B. (1996). Determination of ocular torsion by means of automatic pattern recognition. *IEEE Transactions on Biomedical Engineering*, 43(5), 471-479.

Grundmann, M., Kwatra, V., & Essa, I. (2011, June). Auto-directed video stabilization with robust l1 optimal camera paths. In *Computer Vision and Pattern Recognition (CVPR),* 2011 *IEEE Conference* on (pp. 225-232). IEEE.

Hafed, Z. M., & Clark, J. J. (2002). Microsaccades as an overt measure of covert attention shifts. *Vision research*, 42(22), 2533-2545.

Hartley, R., & Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511811685

Hermens, F. (2015). Dummy eye measurements of microsaccades: Testing the influence of system noise and head movements on microsaccade detection in a popular video-based eye tracker. *Journal of Eye Movement Research*, 8(1).

Herrington, T. M., Masse, N. Y., Hachmeh, K. J., Smith, J. E., Assad, J. A., & Cook, E. P. (2009). The effect of microsaccades on the correlation between neural activity and behavior in middle temporal, ventral intraparietal, and lateral intraparietal areas. *Journal of Neuroscience*, 29(18), 5793-5805.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:*1412.6980.

Ko, H. K., Poletti, M., & Rucci, M. (2010). Microsaccades precisely relocate gaze in a high visual acuity task. *Nature neuroscience*, 13(12), 1549.

Kurzhals, K., Bopp, C. F., Bässler, J., Ebinger, F., & Weiskopf, D. (2014, November). Benchmark data for evaluating visualization and analysis techniques for eye tracking for video stimuli. In *Proceedings of the fifth workshop on beyond time and errors: novel evaluation methods for visualization* (pp. 54-60). ACM.

Laubrock, J., Engbert, R., & Kliegl, R. (2005). Microsaccade dynamics during covert attention. *Vision research*, 45(6), 721-730.

Laurutis, V., Daunys, G., & Zemblys, R. (2010). Quantitative analysis of catch-up saccades executed during two-dimensional smooth pursuit. *Elektronika ir Elektrotechnika*, 98(2), 83-86.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.

Lee, I., Choi, B., & Park, K. S. (2007). Robust measurement of ocular torsion using iterative Lucas–Kanade. *Computer methods and programs in biomedicine*, 85(3), 238-246.

Lord, M. P. (1951). Measurement of binocular eye movements of subjects in the sitting position. *The British journal of ophthalmology*, 35(1), 21.

Lowe, D. G. (2001). Local feature view clustering for 3D object recognition. *In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference* on (Vol. 1, pp. I-I). IEEE.

Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature reviews neuroscience*, 5(3), 229.

Martinez-Conde, S., Macknik, S. L., Troncoso, X. G., & Dyar, T. A. (2006). Microsaccades counteract visual fading during fixation. *Neuron*, 49(2), 297-305.

McCamy, M. B., Otero-Millan, J., Macknik, S. L., Yang, Y., Troncoso, X. G., Baer, S. M., ... & Martinez-Conde, S. (2012). Microsaccadic efficacy and contribution to foveal and peripheral vision. *Journal of Neuroscience*, 32(27), 9194-9204.

Møller, F., Laursen, M., Tygesen, J., & Sjølie, A. (2002). Binocular quantification and characterization of microsaccades. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 240(9), 765-770.

Nachmias, J. (1961). Determiners of the drift of the eye during monocular fixation. *JOSA*, 51(7), 761-766.

Nebehay, G., & Pflugfelder, R. (2014, March). Consensus-based matching and tracking of keypoints for object tracking. In *Applications of Computer Vision (WACV), 2014 IEEE Winter Conference* on (pp. 862-869). IEEE.

Nyström, M., Andersson, R., Niehorster, D. C., & Hooge, I. (2017). Searching for monocular microsaccades–A red Hering of modern eye trackers? *Vision research*, 140, 44-54.

Ong, J. K., & Haslwanter, T. (2010). Measuring torsional eye movements by tracking stable iris features. *Journal of neuroscience methods*, 192(2), 261-267.

Otero-Millan, J., Troncoso, X. G., Macknik, S. L., Serrano-Pedraza, I., & Martinez-Conde, S. (2008). Saccades and microsaccades during visual fixation, exploration, and search: foundations for a common saccadic generator. *Journal of vision*, 8(14), 21-21.

Otero-Millan, J., Castro, J. L. A., Macknik, S. L., & Martinez-Conde, S. (2014). Unsupervised clustering method to detect microsaccades. *Journal of vision*, 14(2), 18-18.

Otero-Millan, J., Serra, A., Leigh, R. J., Troncoso, X. G., Macknik, S. L., & Martinez-Conde, S. (2011). Distinctive features of saccadic intrusions and microsaccades in progressive supranuclear palsy. *Journal of Neuroscience*, 31(12), 4379-4387.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., ... & Lerer, A. (2017). Automatic differentiation in pytorch.

Pizer, S. M., Amburn, E. P., Austin, J. D., Cromartie, R., Geselowitz, A., Greer, T., ... & Zuiderveld, K. (1987). Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, *39*(3), 355-368.

Pekkanen, J., & Lappi, O. (2017). A new and general approach to signal denoising and eye movement classification based on segmented linear regression. *Scientific reports*, 7(1), 17726.

Pelz, J. B. & Witzner Hansen, D. (2017) *International Patent Application No. PCT/US2017/034756*

Pundlik, S. J., Woodard, D. L., & Birchfield, S. T. (2008, June). Non-ideal iris segmentation using graph cuts. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference* on (pp. 1-6). IEEE.

Riggs, L. A., Ratliff, F., Cornsweet, J. C., & Cornsweet, T. N. (1953). The disappearance of steadily fixated visual test objects. *JOSA*, 43(6), 495-501.

Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.

Salvucci, D. D., & Goldberg, J. H. (2000, November). Identifying fixations and saccades in eye-tracking protocols. *In Proceedings of the 2000 symposium on Eye tracking research & applications* (pp. 71-78). ACM.

Serra, A., Liao, K., Martinez-Conde, S., Optican, L. M., & Leigh, R. J. (2008). Suppression of saccadic intrusions in hereditary ataxia by memantine. *Neurology*, 70(10), 810-812.

Shelchkova, N., Rucci, M., & Poletti, M. (2018). *Perceptual enhancements during microsaccade preparation*. Journal of Vision, 18(10), 1278-1278.

Shah, S., & Ross, A. (2009). Iris segmentation using geodesic active contours. *IEEE Transactions on Information Forensics and Security*, 4(4), 824-836.

Tan, C. W., & Kumar, A. (2011, June). Automated segmentation of iris images using visible wavelength face images. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference* on (pp. 9-14). IEEE.

Troncoso, X. G., Macknik, S. L., Otero-Millan, J., & Martinez-Conde, S. (2008). Microsaccades drive illusory motion in the Enigma illusion. *Proceedings of the National Academy of Sciences*.

Yarbus, A. L. (1967). Eye movements during perception of complex objects. In *Eye movements and vision* (pp. 171-211). Springer, Boston, MA.

Zuber, B. L., Stark, L., & Cook, G. (1965). Microsaccades and the velocity-amplitude relationship for saccadic eye movements. *Science*, 150(3702), 1459-1460.