ENABLING SACCADE LATENCY MEASUREMENTS
WITH CONSUMER-GRADE CAMERAS

Hsin-Yu Lai, Gladynel Saavedra-Peña, Charles Sodini, Thomas Heldt, Vivienne Sze

Massachusetts Institute of Technology

ABSTRACT

Eye movements can be affected by a number of neurological, neuromuscular, and neurodegenerative disorders that are important to diagnose and track longitudinally. To enable unobtrusive tracking of disease progression, we tailored and evaluated a set of candidate eye-tracking algorithms to operate on video sequences obtained from an iPhone 6, for accurate and robust determination of the time between the presentation of a visual stimulus and the beginning of the eye movement toward the stimulus (saccade latency). Additionally, we proposed a model-based method to determine the onset of the eye movement and demonstrate that the associated residual normalized root-mean-squared error can be used to automatically flag saccade tracings that should not be included in further analysis. A variant of the iTracker algorithm performs most robustly and results in mean saccade latencies and associated standard deviations on iPhone recordings that are essentially the same as those obtained from simultaneous recordings using a high-end, high-speed camera. Our results suggest that accurate and robust saccade latency determination is feasible using consumer-grade cameras and therefore might enable unobtrusive tracking of neurodegenerative disease progression.

Index Terms— Eye tracking, convolutional neural networks, health monitoring, saccade latency, mobile imaging

1. INTRODUCTION

Neurodegenerative disorders have become increasingly prevalent – in part – due to the expanding elderly population [1]. One example is Alzheimer's Disease, which is the sixth leading cause of death in the United States and one of the nation's costliest health conditions [2]. Diagnosis procedures involve time-consuming neuropsychological tests that have a high retest variability, which makes it difficult to accurately assess the progression of the disease and its response to candidate treatments. Neurodegeneration often manifests in changes in a number of neural pathways, including movement of the eyes [3, 4]. Clinically, such eye movements are typically recorded under controlled conditions using high-cost and special-purpose capital equipment [5, 6]. These restrictions make tracking of disease progression and dynamic assessment of response to treatment difficult. In contrast, low-cost consumer electronic devices are readily available and provide easy access to comparatively high-grade cameras.

In this work, we tailor and evaluate eye-tracking algorithms to enable unobtrusive and repeat measurement of eye-movement patterns on consumer-grade cameras. In particular, we focus on algorithms to measure saccade latency, which is a widely studied reactive eye movement [4, 7]. A saccade is a rapid eye movement between two points of visual fixation, and the saccade latency is defined as the time delay (reaction time) between the appearance of a visual stimulus and when the eye begins to move toward the stimulus [8]. The accuracy of a saccade latency measurement is determined by the accuracy of detecting when the eyes start to move toward the target (saccade onset).

Recently, multiple methods that estimate eye gaze using data collected from mobile devices have been proposed [9]. Among them, convolutional-neural-network-based algorithms [10, 11] have become the state of the art. While these algorithms are tuned to optimize gaze estimation accuracy, this metric does not translate into accuracy of saccade onset detection. Our goal here is to extend a set of candidate eye-tracking algorithms to attain sufficiently precise saccade-latency measurements for clinical applications.

The key contributions of this paper are (1) a candidate algorithm for robust determination of eye movements in video recordings from consumer-grade devices; (2) a model-based approach to saccade onset determination; and (3) automated outlier detection to enable widespread data collection without the need for visual inspection of saccade traces.

2. SACCADE LATENCY MEASUREMENT

There are two main steps in acquiring saccade latencies as shown in Fig. 1: (1) *Eye-tracking* to extract the eye position in an image sequence across time; (2) *Saccade-onset detection* to determine *when* the eye begins to move. In this section, we discuss our experimental set-up, describe the candidate eye-tracking algorithms, and detail our model-based saccade-onset detection scheme that enables automated saccade-latency estimation using consumer-grade cameras.
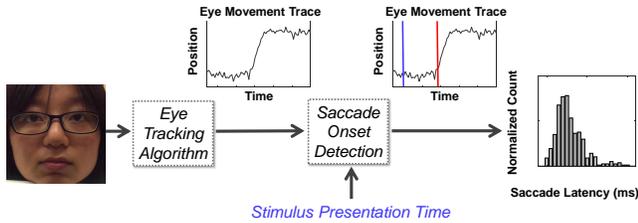
Fig. 1: Pipeline for automated saccade-latency measurement, consisting of eye-tracking and saccade-onset detection. The time difference between the saccade onset (red line) and the stimulus presentation (blue line) is the saccade latency.

2.1. Experimental setup

During our saccade test, subjects are placed in front of a laptop screen with their chin supported. They fix their gaze on a square at the center of a screen. After some time, a visual stimulus appears horizontally to the right or left of the center square. The subjects move their eyes towards the visual stimulus, and back to the center of the screen once the stimulus disappears. This sequence is repeated 40 times. We use the rear-facing camera of a centrally positioned iPhone 6 in slow-motion mode to record the eye movements at 240 frames per second (fps). In a subset of recordings, we simultaneously collect reference videos with a high-speed camera (Phantom v25-11) at 500 fps. Recordings are acquired under standard ambient lighting conditions of a laboratory environment. Details of our data collection and protocol are given in [12].

2.2. Eye-tracking algorithms

Starburst is a feature- and model-based algorithm developed for a head-mounted eye-tracking system [13]. It relies on infrared (IR) illumination to provide a sharp boundary between the pupil and iris (Fig. 2). An initial estimate of the pupil center is used as a seed, and the pupil-iris boundary is detected using gradient-based features along rays that extend radially outward from the seed. RANSAC is used to iteratively fit an ellipse to the detected boundary and arrive at a final estimate of pupil center for each frame [13]. The fixed camera pose relative to the eyes ensures that the eye is always in the same region relative to the camera, which makes algorithm initialization easy across trials. However, the benefits of IR illumination and head-mounting no longer hold when the eye movement is captured with an iPhone camera with a varying pose under natural light.

To address these limitations, we develop **Starburst-phone**. First, we estimate the iris center instead of the pupil center, considering that in visible-spectrum imaging the boundary between the iris and the sclera is often more distinct than the pupil-iris boundary [13] (Fig. 2). With an iPhone, the camera pose can vary, and thus the eye-crop position must be manually determined; by assuming minimal head movement during each test, which lasts under two min-

Fig. 2: Eye images with *(left)* infrared (Fig. from [13]) versus *(right)* natural light.
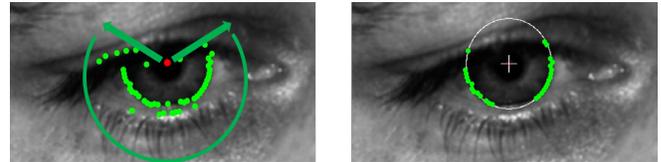
Fig. 3: The Starburst-phone algorithm operating under natural light; *(left)* iris contour detection that avoids the upper eyelid; *(right)* iris model fitting.

utes, the same eye-crop position can be used for all frames. Similarly, the pupil center is also manually initialized in the first frame of each test; however, subsequent frames initialize the pupil center based on the previous frame, which allows for some minor head movement.

Fig. 3 shows how the rays are generated from this initialization point and the gradient along each ray is calculated. We detect the iris contour by choosing the point with the maximum gradient along each ray rather than choosing the point that first exceeds a fixed gradient threshold. Since we are now measuring the boundary between the iris and sclera, the upper eyelid can cause occlusion and the directions of the rays are restricted accordingly. Due to the reduction in the number of rays, we fit a circle model to the iris contour rather than an ellipse. A circle has fewer parameters compared to an ellipse, giving a more stable estimate with fewer feature points. Finally, to adapt to the various lighting conditions, histogram equalization must be selectively applied. Fig. 6 shows an example eye trace using the Starburst-phone algorithm.

iTracker uses a convolutional neural network (CNN) that is trained to determine where a user is looking on a screen (i.e., gaze estimation) based on images taken from a frontal camera of an iPhone or iPad [10]. The inputs of the iTracker include a cropped left eye, a cropped right eye, a cropped face, and a face grid indicating the location of the face within the frame. All the input images have a resolution of 224×224 pixels which means they undergo resizing from the original image; the eye crops are upsampled, while the face crop is downsampled. The architecture of the CNN is shown in Fig. 5. The change in the x-coordinate of the gaze estimation over time corresponds to the eye-movement trace.

While iTracker is designed to operate on a video at 30 frames per second (fps), a temporal resolution below 20 msec is required for clinical applications [14]. Thus for this work, we must use the rear facing camera at 240 fps, which has

Fig. 4: Impact of frame rate on image quality of eye crop; *(left)* 30 fps; *(right)* 240 fps.
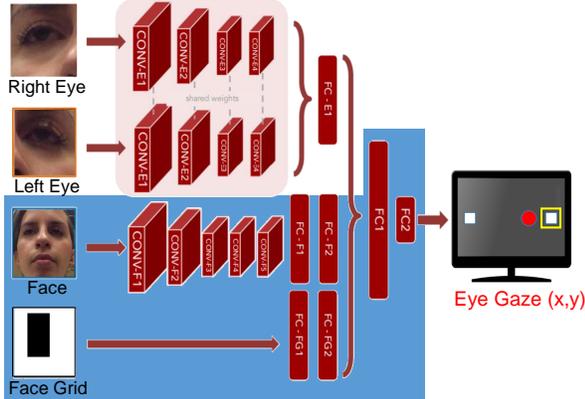
Fig. 5: CNN used by iTracker and iTracker-face. iTracker-face only processes the face layers (highlighted in blue). Figure is modified from [10].

poorer image quality compared to 30 fps due to the reduction in exposure time as shown in Fig. 4.[1] The image quality worsens for the eye crops since they undergo upsampling.

To address the low quality eye-crop image at high frame rate, we propose the **iTracker-face** algorithm, where we only use the face-related convolutional layers as shown in Fig. 5. Although this approach does degrade the accuracy of the gaze estimation as discussed in [10], our objective is to only determine if the gaze changes, and thus it does not significantly affect our results. Fig. 6 shows an example eye trace using the iTracker and iTracker-face algorithms. Note that iTracker-face has a higher signal-to-noise ratio than iTracker and Starburst-phone.

2.3. Saccade-onset detection

Differentiating the eye-position trace has been the method of choice to detect saccade onsets in clinical recordings [15]. Saccade onset is defined as the time at which the eye velocity exceeds 30 degrees/second [15]. To differentiate the eye-position trace, we follow the steps in [16], since [17] suggests it gives physiologically reasonable results. However, differentiation of experimental data is known to amplify high-frequency noise, which motivates our model-based **tanh-fitting** approach to estimate saccade latency. We fit a hyperbolic tangent model to the eye-position trace by minimizing

---

[1]This is not an issue with high-end image sensors such as those found in the Phantom high-speed camera. However, this is one of the reasons the cost for those cameras are in the tens to hundreds of thousand dollars.
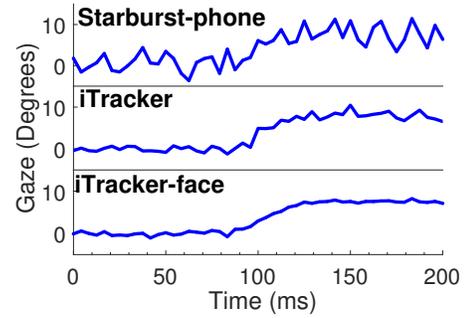
Fig. 6: Sample traces for the three eye tracking algorithms.

**Table 1**: Number of good saccades as determined by visual inspection.

Eye Tracking Algorithm	Glasses *(out of 200)*		Ext. Illumination *(out of 40)*		
	No	Yes	No	Low	Max
Starburst-phone	136	22	17	38	35
iTracker	153	100	29	37	35
iTracker-face	177	181	36	39	35

the root-mean-squared error between the trace and the model fit. The eye velocity is estimated by differentiating the fitted tanh expression, and a saccade onset is determined as the time when the velocity exceeds 30 deg/s. The model-based approach has the added benefit of providing a goodness-of-fit metric on the basis of which the reliability of saccade tracings can be evaluated in an automated manner. The best-fit normalized root-mean-squared error (NRMSE) between the model and the eye-position trace quantifies the residual discrepancy between the two. Unreliable measurements or fits typically have a high NRMSE while reliable measurements have a low NRMSE. Thresholding the NRMSE allows automated rejection of recordings in which the saccade onsets might have been erroneously detected.

3. RESULTS

3.1. Robustness of eye tracking algorithms

To determine the robustness of the candidate eye-tracking algorithms to environmental conditions, we evaluated their performance on video sequences of subjects with and without glasses and under different levels of ambient lighting. An expert annotator reviewed all saccade traces from the three algorithms to determine whether each represents a typical saccade movement and has sufficiently high signal-to-noise to allow saccade-onset determination. Saccades that met these criteria were labeled as good. We compared the number of good saccades across 200 recordings from five subjects in the glasses/no-glasses category and across 40 recordings in one subject for different levels of ambient lighting.
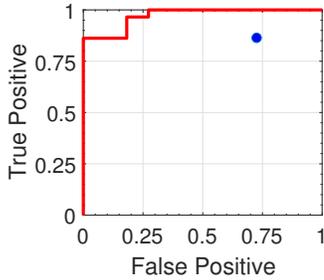
Fig. 7: Performance of the tanh fitting (red curve) and the differentiation method (blue dot) in classifying saccades.

The iTracker-face algorithm achieved the highest number of good saccades across the experimental conditions (Table 1 and Fig. 6). Starburst-phone can mistakenly detect the rim of glasses or hair as part of the iris contour. Under insufficient lighting, it also has difficulty detecting the iris-sclera boundary. Since our model selection in Starburst-phone is based on the maximum number of fitted points under RANSAC, with fewer points detected on the iris contour, the iris fitting will frequently fail. iTracker-face may be less sensitive than iTracker to the details in the eye area of the recordings. While iTracker may need these subtleties to obtain accurate gaze estimation, noise in the eye area during fixation times might obscure the eye movement onset.

3.2. Selection of saccade onset detection method

After identifying iTracker-face as the most robust candidate eye-tracking algorithm for iPhone-based recordings, we compare the differentiation and our proposed tanh-fitting methods for saccade-onset determination on traces produced by iTracker-face. A review of 480 saccade traces revealed that when the traces were sufficiently smooth, the differentiation and tanh methods gave very similar saccade onsets. When traces are noisy, however, the saccade onsets as determined by the differentiation method were prone to error and therefore exhibited significant variation. Since neither erroneous eye movements nor traces with incorrect onsets should be included in saccade-latency determinations and distributions, the differentiation method inevitably requires visual inspection of the saccade onsets. In our experience, the review of 480 saccade traces took about an hour, which makes processing of large volumes of data impractical.

As mentioned in Section 2, an additional benefit of the tanh method is the use of the NRMSE as a natural metric of goodness-of-fit. To evaluate the NRMSE as an automated means to flag bad saccades, we used the expert-annotated saccades for the iTracker-face and tanh fitting as ground truth and swept the NRMSE to generate a receiver operating characteristic (ROC) curve for each recording. A representative ROC curve is shown in Fig. 7, where we also report the true positive/false positive rate attained by the differentiation method (blue dot). We found that for each recording, there ex-
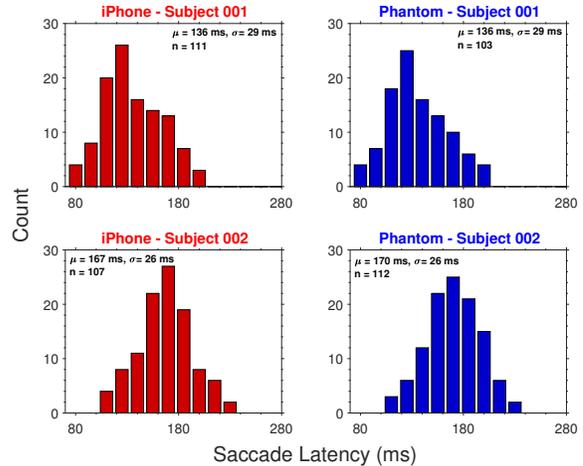
Fig. 8: Example saccade latency distributions obtained from *(left)* iPhone and *(right)* Phantom on two subjects.

ists a wide range of NRMSE thresholds over which the tanh method outperforms the differentiation method by maintaining both a higher true-positive and lower false-positive rate. Moreover, we empirically determined that a NRMSE threshold of 1.5 times the median of the NRMSE across each 40-saccade recording resulted in an average true-positive rate of 0.95 and average false-positive rate of 0.05. The selection of the NRMSE threshold completes the automation of the saccade-latency pipeline in Fig. 1. Analysis and statistical modeling of the saccade-latency distributions obtained with this pipeline are explored in separate work [12].

3.3. Comparison across cameras

To verify that we can attain similar saccade latency statistics using a low-cost, consumer-grade camera (iPhone 6, cost $<$ \$1k, rolling shutter, 720p resolution, 240 fps) as a research-grade camera (Phantom v25-11, cost $\sim$ \$100k, global shutter, 720p resolution, 500 fps), we took simultaneous recordings on two subjects. Fig. 8 shows the resulting saccade latency distributions obtained using the iTracker-face algorithm and tanh-based onset detection. The distributions from both cameras are consistent, with essentially the same mean and standard deviation.

4. CONCLUSION

Our work here demonstrates that iTracker-face, along with the tanh model for saccade-onset determination, is robust to varying recording conditions, allows for automated outlier rejection, and produces saccade latency distributions that are very similar to those obtained from a high-end, high-speed reference camera. This work suggests that tracking of saccade latency can be achieved using the cameras on consumer-grade devices. This contribution paves the way to expanding saccade latency measurements to a broad population for tracking of neurologic and neurodegenerative disease progression.

5. REFERENCES

[1] M.-T. Heemels, "Neurodegenerative diseases," *Nature*, vol. 539, no. 179, 2016.

[2] "Alzheimer's Association," `https://www.alz.org/facts/`, Accessed: 2018-05-21.

[3] O. B. White, J. A. Saint-Cyr, R. D. Tomlinson, and J. A. Sharpe, "Ocular motor deficits in Parkinson's Disease: II. Control of the saccadic and smooth pursuit systems," *Brain*, vol. 106, no. 3, pp. 571–587, 1983.

[4] R. Shafiq-Antonacci, P. Maruff, C. Masters, and J. Currie, "Spectrum of saccade system function in Alzheimer's Disease," *Archives of Neurology*, vol. 60, pp. 1275–1278, 2003.

[5] A.L. Boxer, S. Garbutt, W.W. Seeley, A. Jafari, H.W. Heuer, J. Mirsky, J. Hellmuth, J.Q. Trojanowski, E. Huang, S. DeArmond, J. Neuhaus, and B.L. Miller, "Saccade abnormalities in autopsy-confirmed frontotemporal lobar degeneration and Alzheimer's Disease," *Archives of Neurology*, vol. 69, no. 4, pp. 509–517, 2012.

[6] H.W. Heuer, J.B. Mirsky, E.L. Kong, B.C. Dickerson, B.L. Miller, J.H. Kramer, and A.L. Boxer, "Antisaccade task reflects cortical involvement in mild cognitive impairment," *Neurology*, vol. 81, no. 14, pp. 1235–1243, 2013.

[7] Q. Yang, T. Wang, N. Su, S. Xiao, and Z. Kapoula, "Specific saccade deficits in patients with Alzheimer's Disease at mild to moderate stage and in patients with amnestic mild cognitive impairment," *Age*, vol. 35, no. 4, pp. 1287–1298, 2013.

[8] R.J. Leigh and D.S. Zee, "The neurology of eye movements," chapter 4, p. 174. Oxford University Press, Oxford, 2015.

[9] O. Ferhat and F. Vilariño, "Low cost eye tracking: The current panorama," *Computational Intelligence and Neuroscience*, vol. 3, pp. 1–14, 2016.

[10] K. Krafka, A. Khosla, P. Kellnhofer, H. Kannan, S. Bhandarkar, W. Matusik, and A. Torralba, "Eye tracking for everyone," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2176–2184, 2016.

[11] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "It's written all over your face: Full-face appearance-based gaze estimation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 2299–2308, 2017.

[12] G. Saavedra-Peña, H.-Y. Lai, V. Sze, and T. Heldt, "Determination of saccade latency distributions using video recordings from consumer-grade devices," *Proceedings of the IEEE Engineering in Medicine and Biology Conference (EMBC)*, 2018.

[13] D. Li, D. Winfield, and D.J. Parkhurst, "Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) - Workshops*, pp. 79–87, 2005.

[14] T.J. Crawford, S. Higham, T. Renvoize, J. Patel, M. Dale, A. Suriya, and S. Tetley, "Inhibitory control of saccadic eye movements and cognitive impairment in Alzheimer's Disease," *Biological Psychiatry*, vol. 57, no. 9, pp. 1052–1060, 2005.

[15] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. van de Weijer, "Eye-tracker hardware and its properties," in *Eye Tracking: A Comprehensive Guide to Methods and Measures*, chapter 2, pp. 48–49. Oxford University Press, Oxford, 2011.

[16] M. Nyström and K. Holmqvist, "An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data," *Behavior Research Methods*, vol. 42, no. 1, pp. 188–204, 2010.

[17] L. Larsson, *Event Detection in Eye-Tracking Data for Use in Applications with Dynamic Stimuli*, Ph.D. thesis, Lund University, 2016.