Validation of a Laptop-Webcam-Based Eye Tracking System for Saccade Measurement

Lorenzo Fratini[1], Jaume Pujol[2], José Luis Güell[1] and Clara Mestre[2]

[1] Instituto de Microcirugía Ocular (IMO) Barcelona Grupo Miranza, Carrer de Josep Maria Lladó, 3, 08035 Barcelona, Spain

[2] Centre for Sensors, Instruments and Systems Development, Universitat Politècnica de Catalunya, Rambla Sant Nebridi 10, Terrassa 08222 (Barcelona), Spain

Contact: Lorenzo.fratini@imo.es

In recent decades, video oculography (VOG) has become the most popular eye tracking technique due to its performance, versatility, and minimal intrusiveness. Currently, most commercial video-based eye trackers utilise the pupil-corneal reflection technique, which is based on the detection of the pupil and one or more corneal reflections from an IR light source.[1] VOG has become essential in many fields such as psychology, marketing, or vision science and clinical diagnostics, enabling analysis of eye movements such as saccades, fixations, and smooth pursuits. High-performance commercial eye trackers offer precise measurements; however, their high-cost limits widespread accessibility, particularly for clinical and home-based applications.[2] Recently, deep learning has revolutionised conventional eye-tracking methods based on regressing gaze from the human eye appearance.[3] Deep learning-based eye tracking offers several advantages such as low hardware requirements, increased robustness to head movements, and reduced need for individual calibration procedures. The goal of this study was to compare the performance of a laptop-webcam-based eye tracking system to detect and distinguish between saccades and fixations with a research-grade eye tracker.

Binocular eye movements of 8 healthy adults with normal binocular vision (28.12±3.04 years) were recorded using the Tobii Pro Spectrum and a laptop webcam (Lenovo Thinkbook 14 G7 IML) simultaneously. Head movements were limited with a chinrest. The Tobii Pro Spectrum eye tracker was placed at the recommended distance of 70 cm and captured gaze data at 1200 Hz, while the webcam system was placed at 50 cm from the participants and captured data at 30 Hz. Stimuli were presented on the native Tobii Pro Spectrum screen with a resolution of 1920 x 1080 pixels (52.8 x 29.7 cm). Participants were instructed to follow a red dot (6 mm) on a white background that appeared for 1 to 3 s at the centre of the screen and at eccentricities of ±5, 15, 20 cm. Several Python scripts were developed to synchronize the stimulus presentation with the Tobii Pro Spectrum and laptop's webcam devices. For the webcam, we employed Google's MediaPipe Face Landmarker, a deep learning framework designed for real-time face landmark detection. MediaPipe estimates 3D facial landmarks from live video streams, facilitating the estimation of iris location and eye movements.[4] For the Tobii Pro Spectrum, raw data were exported. First, all data were normalised between [0,1] with respect to the display pixels and then, a Savitsky–Golay filter was used for smoothing raw signals. Eye velocity was computed by differentiation of the position signals, and saccades were detected using an adaptive velocity threshold algorithm. [5]

Figure 1 shows a representative example of horizontal eye position data obtained with the webcam system and the Tobii Pro Spectrum. For this analysis, the amplitudes of saccades made in response to the first 3 stimulus' position changes [from the center of the screen (0.5 in normalised units, norm-units) to +15cm (0.78 norm-units), -15cm (0.22 norm-units) and +5cm (0.59 norm-units)] were included. To study the agreement between the two systems, a Bland-Altman analysis with the data across the eight participants was performed. For the smallest (0.28 norm-units), intermediate (0.37 norm-units) and largest saccadic amplitudes (0.56 norm-units), the mean ± SD of the differences in measured amplitude with the two systems were 0.05 ± 0.04 norm-units, 0.09 ± 0.07 norm-units, and 0.07 ± 0.04 norm-units, respectively. Pooling all saccades together, Bland-Altman analysis showed a bias of +0.017 norm-units with 95% limits of agreement from −0.175 to +0.208 norm-units, indicating the webcam agrees with the Tobii Pro Spectrum within ±0.19 norm-units in amplitude measurement for 95% of individual saccades, and there is no significant bias between the two systems (Figure 2).

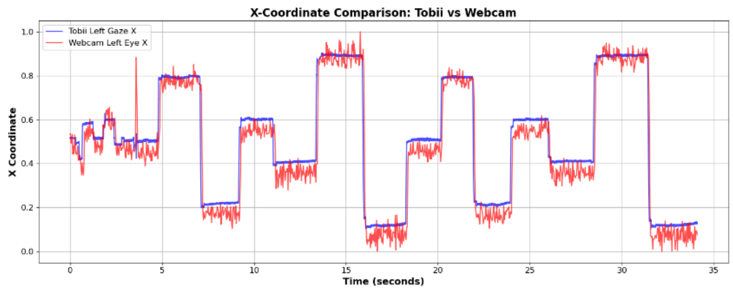

Figure 1: Raw horizontal left eye position data in normalized units obtained with the webcam (red) and the Tobii Pro Spectrum (blue).

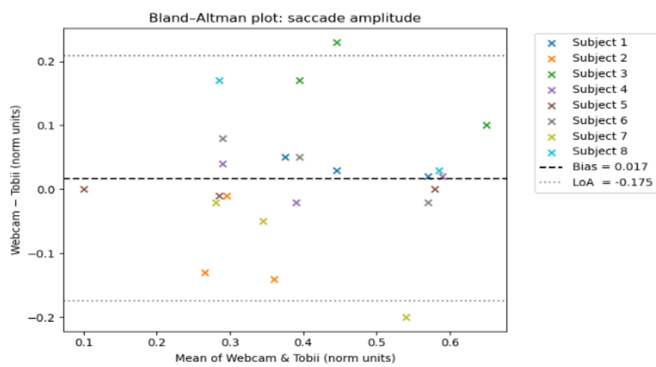

Figure 2: Bland–Altman plot comparing saccadic amplitudes obtained with the webcam and the Tobii Pro Spectrum.

This preliminary study allowed us to first validate this innovative deep learning-based eye tracking system for detecting saccades. Although, on average, the measured amplitudes agreed with those obtained with a high-performance eye tracker, considerable inter-subject variability was found. Future work will deepen the study of the sources of this variability, expand the number of data available, and include different metrics to assess the spatial accuracy, precision and resolution of the system.

ACKNOWLEDGEMENTS: HORIZON–MSCA–2022–DN, *Improving BiomEdical diagnosis through LIGHT-based technologies and machine learning* "BE-LIGHT" (GA nº 101119924 – BE-LIGHT).

---

[1] Mestre, C., Gautier, J., & Pujol, J; Journal *of biomedical optics*, *23*(3), 035001-035001 (2018).

[2] Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & J. Van de Weijer; Oxford University Press (2011).

[3] Cheng, Y., Wang, H., Bao, Y., & Lu, F.; *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2024).

[4] Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., ... & Grundmann, M.; In *Third workshop on computer vision for AR/VR at IEEE computer vision and pattern recognition (CVPR)* (Vol. 2019).

[5] Nyström, M., & Holmqvist, K.; *Behavior research methods*, *42*(1), 188-204 (2010).