Real-Time Webcam-Based Gaze Tracking System for Cognitive and Behavioral Analysis

Elvin Rustamov, Anar Abdullazada, Sadig Latifli, and Rabih Amhaz

[1] UFAZ French-Azerbaijani University, Baku, Azerbaijan
{elvin.rustamov, anar.abdullazada, sadig.latifli}@ufaz.az
https://ufaz.az/

[2] ICube Laboratory, University of Strasbourg, CNRS, France
Icam Strasbourg-Europe, France
rabih.amhaz@icam.fr
https://icube.unistra.fr/, https://www.icam.fr/

Abstract. Our work presents a real-time gaze tracking system that utilizes a standard webcam and the MediaPipe Face Mesh framework to estimate user gaze and attention. The system classifies gaze into a 3x3 screen grid and logs user focus using exponential moving average (EMA) smoothing. Post-session, it visualizes attention heatmaps, enabling cognitive and behavioral analysis. Our goal is to offer a lightweight, privacy-respecting alternative to traditional eye-tracking hardware for real-time user behavior research.

Keywords: Gaze tracking · MediaPipe · Exponential Moving Average · Cognitive analysis · Heatmap visualization

1 Introduction

Understanding user attention and behavior during real-time interactions is a crucial challenge in both education and human-computer interaction (HCI). Traditional eye-tracking hardware, while accurate, tends to be costly and intrusive, limiting its widespread use. In this project we introduce a webcam-based gaze tracking system built with the MediaPipe Face Mesh framework. Our solution is designed to be accessible, real-time, and respectful of user privacy, offering a practical alternative for studying cognitive behavior in classrooms and usability labs.

2 Related Work

Gaze estimation has traditionally relied on hardware-based systems and convolutional neural networks (CNNs). Krafka et al. [1] proposed a crowdsourced dataset and model for mobile eye tracking. Zhang et al. [2] introduced LiteGaze, a lightweight yet accurate model for gaze estimation.

Kuang et al. [3] proposed a 3D deformable eye model to improve estimation under varying lighting and head pose conditions. Patel et al. [4] developed a geometry-based approach using facial landmarks without relying on CNNs.

Webcam-based tracking for screen interaction has been explored by Kim et al. [5], while Lee et al. [6] presented a real-time approach using MediaPipe and CNNs for HCI applications.

3 Methodology

The gaze-tracking pipeline operates through five sequential stages. Firstly, each frame captured by the webcam is processed using the MediaPipe Face Mesh framework, which identifies 468 facial landmarks including detailed points around the eye corners and iris boundaries. In the second stage, we calculate a raw gaze vector by measuring the horizontal and vertical displacement of the iris centroid relative to the corresponding eye corners. Third, these normalized displacement vectors are mapped to one of nine regions on a 3×3 grid (e.g., Top–Left, Center) based on predefined thresholds. To reduce frame-to-frame jitter and improve stability, we then apply an exponential moving average (EMA) filter.

$$G_t = \alpha D_t + (1 - \alpha) G_{t-1}, \quad \alpha = 0.3,$$

where D_t is the raw displacement at time t. Finally, for each frame we log the timestamp, smoothed displacement, and zone label; these logs are later used to generate a heatmap visualizing the user's spatial attention distribution.

3.1 Example Output

An example log entry: `1714671234.123, -0.05, 0.12, Center, Top, Top-Center, 320, 120`
This entry indicates the user was looking at the Top-Center zone, smoothed to coordinates (320, 120). Aggregated logs are used to generate the heatmap.

4 Results

The heatmap shown in Figure 1 visualizes the spatial distribution of gaze zones across a session. It highlights which screen areas received the most attention, enabling inferences about user engagement and cognitive load.

Table 1: Confusion matrix (%)

	Left	Center	Right
Top	0	113	0
Middle	0	950	0
Bottom	0	72	0

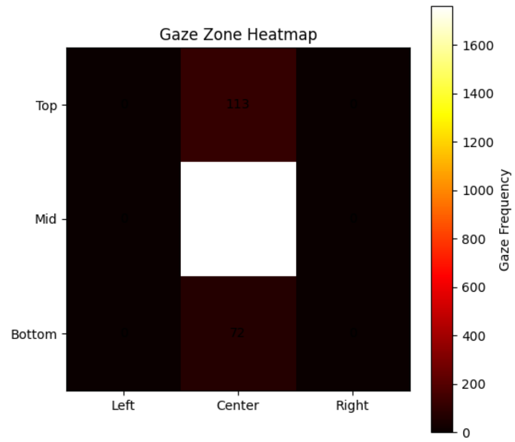

Fig. 1: Heatmap visualizing gaze frequency across screen zones.

These results demonstrate the potential of webcam-based systems in identifying attention patterns, distraction, and fatigue in real-time.

5 Ethical Considerations

Privacy is a core principle of the system. To ensure compliance with ethical standards, no images or video frames are stored at any time and all computations, infrence, and logging are performed locally without any cloud or internet transmission. Only anonymized numeric metrics (timestamps, displacement vectors, and zone labels) are recorded. This design fully complies with GDPR making the system viable for deployment in schools, workplaces, and research labs with strict privacy policies.

6 Conclusion

We presented an affordable, non-intrusive gaze tracking system based on webcam input and MediaPipe Face Mesh. The solution is suitable for cognitive load assessment, attention tracking, and HCI studies. It is especially valuable for environments where hardware-based solutions are impractical. Future improvements may include personalized calibration, higher-resolution zone mapping, and cross-subject learning models for better generalization.

Acknowledgments

This project was supervised by Dr. Rabih Amhaz. We thank the UFAZ faculty and Dr. Rabih Amhaz for their continuous support and feedback during development.

References

1. Krafka, K., Khosla, A., Kellnhofer, P., et al.: Eye Tracking for Everyone. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2176–2184 (2016)
2. Zhang, Y., Chen, X., Wang, Z., Zhou, A.: LiteGaze: A Lightweight and Accurate Gaze Estimation Model. PLOS ONE **18**(5), e0284764 (2023)
3. Kuang, C., Smith, L., Suwajanakorn, S., et al.: Towards an Accurate 3D Deformable Eye Model for Gaze Estimation. arXiv preprint arXiv:2212.01587 (2022)
4. Patel, A., Mehta, R., Chen, L.: Low-cost Geometry-based Eye Gaze Detection using Facial Landmarks. arXiv preprint arXiv:2401.00406 (2024)
5. Kim, D., Choi, S., Park, J.: Webcam-based Gaze Estimation for Computer Screen Interaction. Frontiers in Robotics and AI **11**, 1030187 (2024)
6. Lee, M., Han, S., Yoon, J.: Real-Time Gaze Estimation Using Webcam-Based CNN Models for Human–Computer Interaction. Computers **14**(2), 57 (2024)