PROJECT ABSTRACT

Gaze Tracking using Webcam and Dlib

A webcam gaze tracker that locates eyes with dlib 68-point landmarks, estimates pupil centers by image moments, and maps them to screen coordinates through a buyer-run 9-point calibration.

01. Title

Gaze Tracking using Webcam and Dlib

02. Introduction / Background

Where a person looks is valuable data for attention studies, usability testing and hands-free interface experiments, but commercial eye trackers are specialized hardware. A laptop webcam plus classical vision gets surprisingly far: eye geometry is stable, the pupil is the darkest blob in the eye region, and a short calibration converts pupil positions into screen coordinates. This project builds that pipeline on dlib's 68-point landmark model (Kazemi and Sullivan, 2014): landmarks 36-47 localize both eyes, adaptive thresholding isolates the pupil blob, image moments give its center, and a 9-point on-screen calibration fits a per-user polynomial mapping from pupil vectors to screen coordinates. Because faces, webcams and lighting all differ, accuracy is personal — the included notebook runs the calibration procedure the buyer performs and computes mean angular error over the calibration points, so the report's numbers come from the buyer's own run. The design target is a few degrees of mean angular error at typical laptop distance in good lighting. MPIIGaze (Zhang et al., 2017), the standard in-the-wild appearance-based gaze dataset, is cited as the deep-learning background reference.

03. Problem Statement

Attention and usability experiments need gaze data, but dedicated eye-tracking hardware is expensive and deep gaze networks demand training data and GPUs. The need is a webcam-only gaze tracker built from transparent classical stages — landmark eye localization, pupil-blob center estimation, per-user calibration — whose accuracy is measured by the buyer's own calibration run in degrees of visual angle, with blunt documentation of the lighting, glasses and head-pose conditions under which it breaks.

04. Objectives

The objectives are to localize eye regions with dlib's 68-point landmark model; to estimate pupil centers from thresholded pupil blobs via image moments; to implement a 9-point on-screen calibration procedure fitting a per-user polynomial mapping; to render a real-time gaze-point overlay and log session gaze heatmaps; to report calibration quality with per-point residuals; to gate tracking on head pose so post-calibration movement pauses cleanly; and to evaluate with a notebook procedure computing mean angular error on the buyer's own calibration run.

- Localize eye regions with dlib 68-point landmarks.
- Estimate pupil centers from thresholded blobs via image moments.
- Implement a 9-point calibration with per-user polynomial mapping.
- Render a real-time gaze-point overlay and session heatmaps.
- Report calibration quality with per-point residuals.
- Gate tracking on head pose to handle post-calibration movement.
- Evaluate with a notebook procedure computing mean angular error on the buyer's run.

05. Proposed System / Working

Each webcam frame passes through dlib's face detector and the 68-point landmark predictor; eye regions are cropped from landmarks 36-47 and converted to grayscale. Adaptive thresholding isolates the pupil blob and image moments locate its center; both eyes are averaged into a single gaze vector for robustness. During calibration the user looks at nine on-screen targets while pupil vectors are recorded, and a polynomial mapping is fit to screen coordinates for that user. In tracking mode each frame's pupil vector is mapped through the calibration to a gaze point drawn live on screen, with dwell points logged into a session heatmap. Head-pose gating pauses tracking when the head moves beyond the calibrated pose. The evaluation notebook recomputes the calibration fit and reports mean angular error with residual plots.

- Dlib face detection and 68-point landmark prediction.
- Eye-region cropping with adaptive pupil thresholding.
- Image-moment pupil-center estimation, both eyes averaged.
- 9-point calibration with per-user polynomial mapping.
- Live gaze-point overlay and session heatmap logging.
- Head-pose gating and notebook evaluation with residual plots.

06. Technologies / Components Used

The system uses Python 3.10 with dlib for face detection and 68-point landmarks, OpenCV for frame capture, eye cropping and thresholding, NumPy for pupil geometry and mapping math, scikit-learn for the polynomial calibration fit, a Jupyter notebook for the calibration procedure and error analysis, matplotlib for calibration-grid and residual plots, and a Tkinter overlay app for calibration targets and the live gaze point.

- Python 3.10, dlib (face detection, 68-point landmarks).
- OpenCV (frame capture, eye cropping, thresholding).
- NumPy (pupil geometry and mapping math).
- scikit-learn (polynomial calibration fit).
- Jupyter notebook (calibration procedure and error analysis).
- Matplotlib (calibration grid and residual plots).
- Tkinter overlay app (calibration targets and live gaze point).

07. Key Features

Key features include dlib 68-point eye-region localization, pupil-center estimation via thresholded blobs and image moments, a 9-point on-screen calibration with per-user mapping, a real-time gaze-point overlay, session gaze-heatmap logging, a calibration-quality report with per-point residuals, an evaluation notebook computing mean angular error on the buyer's run, and head-pose gating that pauses tracking on large head movement.

- Dlib 68-point eye localization.
- Pupil-center estimation by image moments.
- 9-point calibration with per-user mapping.
- Real-time gaze-point overlay.
- Session gaze-heatmap logging.
- Calibration-quality report with residuals.
- Notebook evaluation: mean angular error on the buyer's run.
- Head-pose gating for post-calibration movement.

08. Applications / Use Cases

Webcam gaze tracking applies to usability-testing attention studies, hands-free interface experiments, reading-behavior research prototypes, and presentation attention demos. It also demonstrates landmark-based vision, calibration mathematics and honest error reporting as viva-ready concepts — always framed as an educational prototype, never a medical or assistive device.

- Usability-testing and attention studies.
- Hands-free interface experiments.
- Reading-behavior research prototypes.
- Teaching landmark vision and calibration math.
- Demonstrating honest error reporting in a viva.

09. Results & Scope

The expected outcome is a working gaze tracker whose accuracy is measured, not asserted: the included notebook runs the buyer's own 9-point calibration and computes mean angular error in degrees of visual angle, with residual plots showing per-point fit quality. The design target is a few degrees of mean angular error at typical laptop distance in good lighting; the buyer's run settles the actual number. Scope covers a single user, fixed seating position, and frontal webcam views in adequate lighting; documented limits include lighting as the dominant failure mode, calibration invalidated by head movement, glasses glare and eyelid occlusion, and webcam resolution bounds. This is an educational attention-tracking prototype — it must not be used for diagnosis, accessibility-critical control or any safety-relevant purpose. Future scope includes head-pose-compensated mapping, blink detection, and multi-user calibration profiles.

- Measured outcome: mean angular error computed by the notebook on the buyer's own calibration run.
- Residual plots documenting per-point fit quality.
- Design target: a few degrees of mean angular error at laptop distance in good lighting.
- Scope: single user, fixed seating, frontal webcam, adequate lighting.
- Documented limits: lighting, head movement, glasses glare, resolution.
- Honest framing: educational prototype, never medical or safety-relevant.
- Future scope: head-pose compensation, blink detection, multi-user profiles.

10. Conclusion

Dlib landmarks plus pupil-blob geometry plus a 9-point calibration turn an ordinary webcam into a personal gaze tracker, and this project delivers it as a complete auditable system: eye localization, pupil estimation, calibration, live overlay, heatmap logging, and a notebook that measures mean angular error on the buyer's own run. Built to order, it ships with source code, the calibration and evaluation notebook, report, presentation, and viva preparation material.

11. References

The project cites the dlib landmark paper, the MPIIGaze dataset paper as deep-learning background, and the dlib library reference.

- V. Kazemi and J. Sullivan, 'One Millisecond Face Alignment with an Ensemble of Regression Trees,' Proc. CVPR, 2014.
- X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, 'It's Written All Over Your Face: Full-Face Appearance-Based Gaze Estimation,' Proc. CVPR, 2017.
- D. E. King, 'Dlib-ml: A Machine Learning Toolkit,' J. Machine Learning Research, 2009.