Human Interaction and Emerging Technologies (IHIET 2025), Vol. 197, 2025, 293–303
https://doi.org/10.54941/ahfe1006722

Evaluating Automated Gaze Mapping Across Laboratory and Field Study Settings

Celina Vetter[1], Rebecca Nauli[1], Ruth Häusler Hermann[1], and Maarten Uijt de Haag[2]

[1]Zurich University of Applied Sciences, Winterthur, Switzerland
[2]Technical University Berlin, Berlin, Germany

ABSTRACT

Eye Tracking (ET) is used across various fields to study visual attention, cognitive load, and decision-making by measuring where individuals look and how long they focus on specific elements. A major challenge in processing ET data is mapping gaze points to dynamic Areas of Interest (AOIs) while accounting for data variability and head movements. This study aims to systematically compare and evaluate methods for automated gaze-to-AOI mapping across three experimental conditions with varying levels of control, in order to improve efficiency and accuracy in gaze analysis. An analytical software based on ArUco-markers was developed to automate gaze mapping to AOIs with three different methods: (1) marker-based mapping, and homography-based mapping implemented using either (2) manually specified reference points or (3) automated feature detection algorithms. All three methods were compared against two baselines: manual gaze mapping and assisted mapping using a commercial ET software. Overall, the results indicate that the performance of automated gaze-to-AOI mapping methods is highly sensitive to experimental context and the specific configuration of AOIs. Under laboratory conditions, automated gaze mapping methods achieved accuracy (97%) and F1-scores (97%) comparable to manual mapping. In complex field study settings, the performance varied, and accuracy dropped (14% to 77%) due to varying conditions regarding sudden transition in lighting and real-time dynamics of the situation. The manual reference point-based method demonstrated consistently high accuracy across all experimental conditions. Only manual mapping remained highly accurate in both field study conditions but required considerably more processing time. Future work will focus on improving the robustness of the proposed method in dynamic environments through adaptive reference image selection. This enhancement is expected to increase the accuracy of gaze-to-AOI mapping and enable real-time monitoring of visual attention in complex, safety-critical contexts. Such advancements will support the development of resilient, adaptive human-machine systems capable of dynamically responding to operator conditions, thereby reducing the risk of human error and improving overall performance.

Keywords: Eye tracking data processing, Gaze point mapping, Dynamic areas of interest, Automatic gaze analysis methods

INTRODUCTION

In an increasingly automated and data-driven world, understanding how humans perceive and interact with technical systems and their environment is crucial for fostering effective human-machine collaboration, ensuring that machines complement human capabilities and enhance overall performance. Eye tracking (ET) is a well-established method to gain insights into human cognitive processes used across aviation, automotive systems, industrial control rooms, or consumer product testing. By capturing where, for how long and in what sequence a person focuses their visual attention, ET allows to understand cognitive workload, analyse decision-making, and assess human performance in both laboratory and field study environments.

ET has evolved from a niche method to a widely used technology (Valliappan et al., 2020). As ET becomes more accessible, its applications are expanding across various industries (Martinez-Marquez et al., 2021; Hebbar et al., 2022; Xu et al., 2018; Tahri Sqalli et al., 2023; Kim & Lee, 2021). Industries are increasingly recognizing the need to efficiently and accurately analyse gaze data in relation to specific Areas of Interest (AOIs) within dynamic visual environments.

Mapping raw gaze data to AOIs is an error-prone and time-intensive task when done manually, especially in dynamic scenes with moving AOIs or unfixed head positions. These challenges highlight the need for automated, robust, and scalable mapping methods (Justinussen, 2025) (Kopácsi et al., 2023). Approaches that address mapping limitations have been developed. Machine learning approaches like Computational Gaze-Object Mapping (cGOM) (Wolf et al., 2018), and AutoAOIs scale well to dynamic AOIs (Justinussen, 2025). Methods such as Interactive Machine learning for Efficient Tracking of AOIs (IMETA) blend manual labelling with Machine Learning to cut annotation time (Kopácsi et al., 2023). Marker-based systems (e.g., ArUco) offer accuracy in static, laboratory scenes (Barz & Sonntag, 2021; Bykowski & Kupiński, 2018). For naturalistic settings, 3D reconstruction methods like MAP3D enable marker-free gaze projection onto spatial models, though at high computational cost (Stein et al., 2023; Bykowski & Kupiński, 2018). An overview of strengths and weaknesses of these methods is provided in Table 1.

Table 1: Comparison of different gaze mapping methods.

Method	Strengths	Weaknesses
Manual mapping	High accuracy in laboratory settings	Time-consuming, error-prone, not scalable
Manual mapping + head tracking	Improved accuracy	Expensive hardware, calibration and space requirements
Marker-based systems (e.g. ArUco)	Fast, stable reference in static scenes	Visually intrusive, lab-bound
Machine learning (e.g. cGOM)	Scalable, handles dynamic AOIs	Requires labelled training data

Continued

Table 1: Continued

Method	Strengths	Weaknesses
Semi-automatic (e.g. IMETA)	Reduces manual effort, balances control	Needs initial labelling, performance depends on input quality
AutoAOIs	Tracks moving AOIs automatically	Object detection errors possible
3D Reconstruction (e.g. MAP3D)	Marker-free, works in naturalistic settings	Computationally intensive, complex setup

Despite recent advances, selecting an appropriate gaze mapping method remains challenging. Performance often depends on data quality and the research design. This underscores the need for systematic evaluation to support an informed method selection.

The present study compares and evaluates three different methods for mapping gaze points to AOIs against two baselines. This study considers: (1) ArUco marker-based mapping anchoring AOIs in the physical setup, and homography-based mapping that uses either (2) manual-defined reference points or (3) feature detection to transform gaze data into the coordinate space of the AOIs. The baselines include both manual gaze mapping and assisted mapping using the commercial software Tobii Pro Lab. The methods are tested in both laboratory and field study settings to compare their accuracy and robustness.

The research question is "Can automated gaze mapping methods match the precision achieved by manual mapping while reducing processing time in complex environments?" with the following hypotheses:

1. H1: Automated gaze mapping methods (marker-based & homography-based) will achieve accuracy comparable to manual mapping under laboratory conditions.
2. H2: The accuracy of automated gaze mapping will decrease in field study environments but will remain at or above 90% of the level of accuracy reached by manual mapping.
3. H3: Automated mapping methods will considerably reduce processing time (setup and execution time) compared to manual mapping.
4. H4: Among the automated methods, the marker-based method will demonstrate consistently the highest overall accuracy across all settings.

METHODS

This chapter outlines the core functionality, and evaluation of the gaze mapping analytical software, including its methods and test settings.

The goal of the analytical software is to determine exactly when and where a person was looking by combining ET data with spatial markers visible in the environment. The analytical software implements three methods to choose from that spatially map gaze points, determine whether they fall within predefined AOIs, and generate both visual and quantitative outputs to support interpretation:

1. Marker-Based Mapping: In this method, AOIs are defined directly within the video frame by using ArUco markers detected dynamically in each frame.
2. Feature Detection-Based Homography Mapping: This method uses automatic feature detection to align each video frame with a predefined reference image of the setting. A homography matrix is computed per frame to transform gaze coordinates from the video into the reference image coordinate system. This mapping method is useful when ArUco markers are not always visible due to head movements.
3. Manual Reference Point-Based Homography Mapping: In this method, the user manually defines reference points on both the reference image and corresponding points in the video frames. These points are used to compute a homography transformation that maps gaze coordinates from the video to the reference image. ArUco markers can be used as known reference points to facilitate accurate correspondence. This method is appropriate when the setting has few distinct features.

Analytical Software Architecture

First the processing steps for gaze data are described in chronological order referred by numbers. Before marker detection, the analytical software runs an ET calibration to align gaze data with relevant elements of the setting. Offsets between gaze points and targets are visually verified. The calibration uses a Charuco board to estimate the ET camera's intrinsic parameters (camera matrix and distortion). Calibration images are processed to detect marker and Charuco corners, which are used by OpenCV to compute the camera model. Parameters are saved in undistorted, gaze-aligned frames. Calibration is needed once per setting.

The overall processing pipeline of the analytical software is depicted in Figure 1, with the two gaze mapping methods visualised in orange and purple further detailed in Figure 2. User-defined inputs (in blue, on left side in Figure 1) and outputs (in green, on right side on Figure 1) are further described in the next chapter.

Following initial gaze data preparation (3), the first key processing step is the ArUco marker detection (5) using OpenCV's ArUco libraries, which establishes a spatial reference by identifying markers frame by frame in the video.

Once markers are detected, the system merges gaze and marker data (7), synchronising data streams based on timestamps. If gyroscopic data is available (8), it stabilises marker detection (9). When head movement, measured by gyroscopic data, stays below a threshold, ArUco marker detections are extrapolated, even if markers temporarily disappear.

The aligned data can then be processed by two gaze mapping methods (10): marker-based (10a) or homography-based (10b). In the marker-based mapping (10a), AOIs are calculated dynamically using ArUco marker positions in each video frame. AOI hits are calculated, and the analytical software overlays AOI outlines directly onto the ET video to support visual inspection of gaze behaviour.

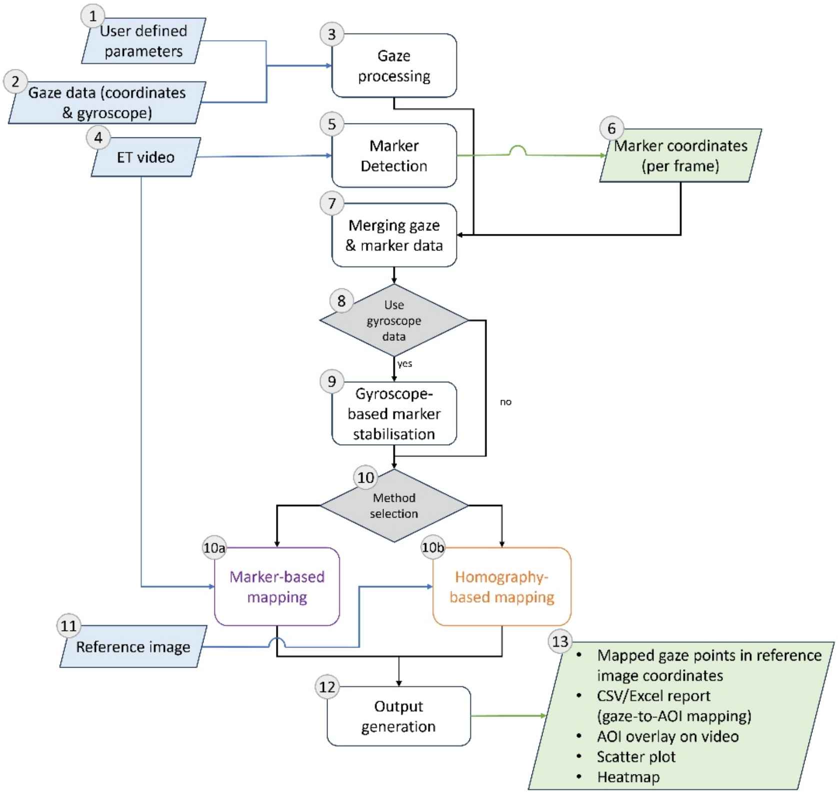

Figure 1: Overview of the gaze mapping tool architecture.

In contrast, the homography-based mapping (10b) transforms gaze points from video coordinates to a static reference image using a homography matrix. This transformation can be based either on manually defined reference points or on feature detection using Scale-Invariant Feature Transform (SIFT) and Fast Library for Approximate Nearest Neighbors (FLANN) (OpenCV, 2025).

In more detail, feature points are extracted from both the video frame and reference image, with FLANN using a KD-tree (5 trees, 20 checks) for efficient nearest-neighbour search in the 128-dimensional SIFT descriptor space. Matches are found via k-nearest neighbours ($k = 2$) and ranked by Euclidean distance. AOI hits are subsequently determined by mapping gaze data onto the reference image.

Input and Output

The analytical software requires four main input data, as shown in Figure 1, depending on the processing method: user-defined parameters (1), gaze coordinates in pixels (optional including gyroscopic data) (2), ET video (4) and a reference image (11). The video from the eye tracker is paired with gaze data frame by frame. The analytical software is hardware-agnostic, given the gaze coordinates are in the same 2D coordinate system as the video frames.

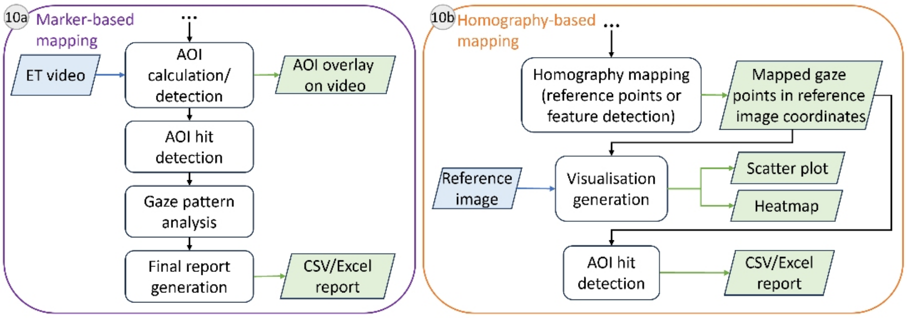

Figure 2: Detailed visualisation of the two gaze mapping methods mentioned in Figure 1.

The user defined parameters (1) include paths to the input files (gaze data & ET video), flags to enable or disable certain steps like heatmap generation, homography matching, or gyroscope integration and the definitions of AOIs.

Defining AOIs necessitates precise specification based on the spatial configuration of marker positions. Each rectangular AOI is anchored to a specific ArUco marker, identified by its unique marker ID and one of its four corners, which serves as the reference point for AOI placement. In manual reference point-based homography mapping, it is essential to specify the positions of these markers in pixel coordinates within the reference image. Based on this anchor, AOIs are defined by a name, a fixed size (width, height), and an optional rotation. To allow flexible positioning, AOIs can also be placed using an offset relative to the anchor, enabling precise specification of AOIs. All parameters are specified in a Python-based configuration file.

Outputs (12 & 13) include several visualisations such as gaze plots, heatmaps and AOI overlay on video. The analytical software compiles a final report containing time series of AOI hits, number of fixations and mean fixation dwell times per AOI. This report can be saved as a CSV or Excel file for further analysis.

Validation Approach

The validation of automated gaze mapping is conducted with data from a laboratory setting representing acceptable to ideal conditions for ET, and two datasets from a field study incorporating typical challenges to ET such as inconsistent lighting, and marker detectability. The three data sets are described in Table 2.

Table 2: Description of validation data sets.

	Screen Recording (Laboratory)	Flight Simulator (Field Study)	Train Cockpit (Field Study)
Video duration [sec]	27	2030	1672
# of frames [-]	682	50751	41814
Sampling rate [Hz]	25	25	25
Gaze samples [%]	98	92	84

The laboratory setting called “screen recording” represents an ideal setting for automated gaze mapping: ArUco markers remained constantly visible, lighting conditions were optimal, and gaze input was simulated using the mouse cursor for better control. The first field study setting was a low-fidelity flight simulator and the second was a train driver’s cab. The environment in both field study settings exhibited high dynamic range lighting, with high luminance contrast between the bright exterior view and the darker interior of the cockpit/cab. Also, the tasks involved extensive head movements shifting visual attention.

Validation involved comparing three automated gaze mapping methods against two baselines: (i) manual mapping, regarded as the ground truth, and (ii) automated gaze mapping provided by Tobii Pro Lab (Tobii Pro Lab, 2025). In the first, screen recording conducted under optimal conditions, the baseline consisted of the actual recorded data, as both the mouse cursor (used to simulate gaze) and ArUco markers were always visible and accurately trackable.

Comparison of the gaze mapping methods is based on performance metrics for time efficiency, accuracy of AOI hits (the proportion of correct predictions), precision (the proportion of true positive hits among all predicted positive hits), recall (the proportion of true positive hits among all actual positive hits), and F1-score (the harmonic mean of precision and recall). A server with 8×3.2GHz CPU cores, 16 GB of RAM and no GPU was used for testing.

RESULTS

Computational Efficiency

To evaluate the computational efficiency of the analytical software, the setup and execution time for each processing step was measured across all datasets. Defining input parameters took 15 minutes for both homography methods, with an extra 15 minutes for manual reference point-based mapping using 12 markers. These inputs are needed only once per setting.

The total computational time for all methods is shown in Table 3, with homography matrix calculation times in brackets. Table 3 indicates considerably higher computational costs for feature detection compared to the manual reference point-based mapping. Most processing steps required similar durations, except the homography matrix calculation. Manual reference point-based mapping required between 75–97% less time than manual mapping depending on the setting, while feature detection required 191% (flight simulator) and 173% (train driver’s cab) more time than the manual gaze mapping.

Table 3: Total computational time (incl. time for homography matrix calculation in brackets).

Method\Setting	Screen Recording	Flight Simulator	Train Driver’s Cab
Baseline	10.4 min	119.3 min	104.3 min
Manual reference point	0.3 min (0.02 min)	29.8 min (0.5 min)	17.1 min (0.6 min)

Continued

Table 3: Continued

Method\Setting	Screen Recording	Flight Simulator	Train Driver's Cab
Feature detection	3.7 min (3.6 min)	346.9 min (334.8 min)	284.9 min (278.6 min)

Accuracy Comparison

Table 4 shows the accuracy parameters calculated for mapping quality of the gaze mapping methods in laboratory and field study conditions.

Table 4: Performance metrics for gaze-to-AOI mapping across methods and datasets.

Setting	Method	Accuracy [%]	Precision [%]	Recall [%]	F1-Score [%]
Screen recording	Tobii Pro Lab (manual/assisted)	93/100[1]	95/100[1]	92/100[1]	94/100[1]
	ArUco-marker	97	96	98	97
	feature detection	93	93	95	94
	manual reference points	90	94	89	91
Flight simulator	Tobii Pro Lab (manual/assisted)	100[1]/32	100[1]/94	100[1]/13	100[1]/23
	ArUco-marker	70	85	72	78
	feature detection	72	77	90	83
	manual reference points	72	77	90	83
Train driver's cab	Tobii Pro Lab (manual/assisted)	100[1]/NA[2]	100[1]/NA[2]	100[1]/NA[2]	100[1]/NA[2]
	ArUco-marker	38	76	14	24
	feature detection	14	24	16	19
	manual reference points	77	82	91	86

[1] 100% indicates it was the ground truth.
[2] NA indicates no valid gaze points were detected, preventing metric calculation.

Overall results indicate that the analytical software's performance varied across mapping method and setting. Manual reference point-based homography generally showed consistent high accuracy (72–90%) and F1-scores (83–91%) where reference points were well-defined across all three datasets. In contrast, feature detection-based homographic mapping varied in accuracy, reflecting differences in setting complexity affecting the fit of reference image. Tobii Pro Lab, performed well in laboratory environments but failed in complex field study settings.

Accuracy was highest for all gaze mapping methods in the laboratory setting "screen recording" and lowest in the field study condition (train driver's cab), where the feature-based method showed poor precision and recall due to inconsistent key point matching.

DISCUSSION

The results demonstrate that performance of the mapping methods highly depends on experimental conditions. None of the three methods for automated gaze mapping (marker-based, feature detection-based, manual reference point-based mapping) showed sufficiently high accuracy across all three experimental conditions. Results indicate that mapping accuracy improved under more controlled conditions, with the highest accuracy (>90%) across all three mapping methods. This can be attributed to the highly controlled experimental environment, including consistent lighting, clear marker visibility and mouse-simulated gaze data. This confirms the analytical software's reliability under optimal conditions.

In more complex field study settings like the flight simulator and train driver's cab automated gaze mapping faces multifaceted challenges such as lighting variability, motion blur, and moving backgrounds, which degraded homography accuracy, especially for feature detection. These findings highlight how sensitive homography-based methods are to visual dissimilarities of the reference image and the video content. Marker-based mapping, while more flexible, was sensitive to lighting conditions, as poor illumination or glare often disrupted marker detection resulting in reduced mapping reliability.

The performance of the analytical software partly depends on the accuracy of ET calibration: small inaccuracies are generally tolerated, except near AOI boundaries. The size of the AOI influenced mapping accuracy: small AOIs increased mapping loss, while large AOIs risked inclusion of irrelevant fixations.

Setup duration was similar across methods (15–30 minutes to define reference images, parameters, and AOIs). Feature detection reduced manual input but incurred higher computational costs and was highly sensitive to uncontrolled conditions. Therefore, selecting a mapping method tailored to the experimental setting, rather than applying all available methods indiscriminately, can considerably reduce overall processing time.

Efficiency can be further improved by restricting the analysis to relevant video segments, whereby reducing processing time without losing valuable data. Future improvements need to focus on selecting reference image adaptively to assure gaze-mapping accuracy in dynamic environments. Table 5 provides a summary of the results in relation to the hypotheses under investigation.

Table 5: Overview of hypothesis outcomes.

Hypothesis	Outcome	Explanation
H1 (accuracy in laboratory settings)	Supported	Under controlled conditions, all automated methods achieved high accuracy comparable to the manual baseline.
H2 (accuracy in field study settings)	Not supported	Automated methods did not reach 90% accuracy in more complex environments.
H3 (low processing time)	Supported	Automated methods reduced processing time, though runtime was increased for feature detection.
H4 (marker-based superior)	Not supported	Marker-based mapping was generally accurate, but in some cases, manual homography mapping outperformed.

CONCLUSION

Automated gaze mapping methods have the potential to considerably reduce time and effort needed to analyse visual attention in complex environments, as long as they maintain adequate accuracy.

This study compared three methods for automated gaze mapping (marker-based mapping, feature detection-based homography mapping, and manual reference point-based homography mapping) with manual mapping and commercial mapping software. Mapping accuracy was evaluated across three experimental conditions with varying complexity including a laboratory condition ("screen recording"), and two field study conditions ("flight simulator", and "train driver's cab").

The results showed that accuracy of the methods for automated mapping varied depending on environmental factors such as lighting conditions, scene dynamics, and calibration accuracy of eye-tracking. While automated methods achieved high accuracy in the laboratory condition, their performance decreased with more challenging lighting conditions in field study settings. The manual reference point-based method showed the most consistent results across the three test settings. In terms of computational efficiency, automated methods considerably reduced the need for manual input, although feature-based homography induced higher processing times.

The results of the study highlight the trade-offs between accuracy, automation of labour-intensive manual gaze-to-AOI mapping processes, and runtime, as well as the importance of AOI definition and fit of the reference image. The findings suggest that automated mapping is a viable alternative to manual processing in laboratory conditions.

Future improvements will focus on optimising processing time, extending visualisation capabilities, and increasing robustness to environmental variability.

Outlook

This work lays a foundation for enhanced resilience in human-machine collaboration by enabling accurate, real-time monitoring of visual attention that is indicative for user situation awareness and intent, as well as task execution through robust, automated gaze-to-AOI mapping. Future implementations in human-machine systems could integrate real-time gaze analysis to enhance safety by identifying instances when critical elements are overlooked, thereby triggering timely, adaptive system responses. Additionally, future work will explore adapting the analytical software to process shorter time windows with dynamically selected reference images to improve gaze mapping accuracy in visually changing environments. In future, the analytical software's architecture may integrate physiological and contextual data streams, further supporting the development of adaptive, attention-aware assistant systems in safety-critical domains such as aviation, rail, and healthcare.

REFERENCES

Barz, M. & Sonntag, D., 2021. Automatic Visual Attention Detection for Mobile Eye Tracking Using Pre-Trained Computer Vision Models and Human Gaze. Sensors, 21(12)(4143).

Bykowski, A. & Kupiński, S., 2018. Feature matching and ArUco markers application in mobile eye tracking studies. Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), pp. 255–260.

Hebbar, P. A., Pashilkar, A. A. & Biswas, P., 2022. Using eye tracking system for aircraft design – a flight simulator study. Aviation, Volume 26(1), pp. 11–21.

Justinussen, J., 2025. Gaze-Mapping vs AutoAOIs – How To Choose The Best Eye Tracking Analysis Tool. [Online] Available at: https://imotions.com/blog/insights/research-insights/gaze-mapping-vs-autoaoi/.

Kim, N. & Lee, H., 2021. Assessing Consumer Attention and Arousal Using Eye-Tracking Technology in Virtual Retail Environment. Frontiers in Psychology, 12(665658).

Kopácsi, L., Barz, M., Shahzad Bhatti, O. & Sonntag, D., 2023. IMETA: An interactive mobile eye tracking annotation method for semi-automatic fixation-to-AOI mapping. In Companion Proceedings of the 28th International Conference on Intelligent User Interfaces, pp. 33–36.

Martinez-Marquez, D. et al., 2021. Application of Eye Tracking Technology in Aviation, Maritime, and Construction Industries: A Systematic Review. Sensors, 21(13)(4289).

OpenCV, 2025. Feature Detection and Description. [Online] Available at: https://docs.opencv.org/4.x/db/d27/tutorial_py_table_of_contents_feature2d.html.

Stein, I., Jossberger, H. & Gruber, H., 2023. MAP3D: An explorative approach for automatic mapping of real-world eye-tracking data on a virtual 3D model. Journal of Eye Movement Research, 15(3)(16910).

Tahri Sqalli, M. et al., 2023. Eye tracking technology in medical practice: A perspective on its diverse applications. Frontiers in Medical Technology, Volume 5.

Tobii Pro Lab, 2025. Behavior research software with eye tracking. [Online] Available at: https://www.tobii.com/products/software/behavior-research-software/tobii-pro-lab.

Valliappan, N. et al., 2020. Accelerating eye movement research via accurate and affordable smartphone eye tracking. Nature Communications, 11(4553).

Wolf, J. et al., 2018. Automating Areas of Interest Analysis in Mobile Eye Tracking Experiments Based on Machine Learning. Journal of Eye Movement Research, Volume 11(6), pp. 1–11.

Xu, J., Min, J. & Hu, J., 2018. Real-time eye tracking for the assessment of driver fatigue. Healthcare technology Letters, Volume 5(2), pp. 54–58.