A Low-Power Distance-Based Approach to Gaze Classification in Pervasive Smart-Eyewear Eye Tracking

Carlo Pezzoli
DEIB
Politecnico di Milano
Milano, Italy
carlo.pezzoli@polimi.it

Flavia Nicotri
DEIB
Politecnico di Milano
Milano, Italy
flavia.nicotri@mail.polimi.it

Marco Brando Mario Paracchini
DEIB
Politecnico di Milano
Milano, Italy
marcobrando.paracchini@polimi.it

Luca Francesco Raduzzi
DEIB
Politecnico di Milano
Milano, Italy
lucafrancesco.raduzzi@polimi.it

Daniele Bani
EssilorLuxottica Smart Eyewear Lab
EssilorLuxottica
Milano, Italy
daniele.bani@luxottica.com

Giulio Marano
EssilorLuxottica Smart Eyewear Lab
EssilorLuxottica
Milano, Italy
giulio.marano@luxottica.com

Luca Merigo
EssilorLuxottica Smart Eyewear Lab
EssilorLuxottica
Milan, Italy
luca.merigo@luxottica.com

Marco Marcon
DEIB
Politecnico di Milano
Milano, Italy
marco.marcon@polimi.it

Abstract

This work presents a low-power embedded eye-tracking system based on Photosensor Oculography (PSOG) and evaluates lightweight calibration-driven distance-based classifiers for discrete gaze-zone estimation in smart eyewear. The proposed prototype integrates infrared LEDs and photodiodes within a commercial frame and uses sequential illumination to generate 32-dimensional feature vectors at 1 kHz with minimal computational cost. Distance-based classifiers, including Euclidean, standardized Euclidean, and Mahalanobis metrics, were first evaluated using a humanoid robot with human-like ocular geometry under controlled conditions. The selected methods were then extensively evaluated on a human dataset comprising 29 subjects under varying illumination conditions. Results show that the Mahalanobis distance achieves the highest accuracy, while dimensionality reduction improves robustness under changing lighting. A PCA-Mahalanobis configuration provided the best accuracy-complexity trade-off, while stronger feature reduction combined with Euclidean distance enabled competitive performance at very low computational cost, supporting deployment on both high-performance and embedded smart eyewear platforms.

CCS Concepts

• **Computing methodologies** → **Classification and regression trees**; • **Human-centered computing** → *Ubiquitous and mobile computing systems and tools*; • **Software and its engineering** → Embedded software.

ETRA '26, Marrakesh, Morocco

ACM ISBN 979-8-4007-2519-7/26/06
https://doi.org/10.1145/3797246.3806216

ACM Reference Format:
Carlo Pezzoli, Flavia Nicotri, Marco Brando Mario Paracchini, Luca Francesco Raduzzi, Daniele Bani, Giulio Marano, Luca Merigo, and Marco Marcon. 2026. A Low-Power Distance-Based Approach to Gaze Classification in Pervasive Smart-Eyewear Eye Tracking. In *2026 Symposium on Eye Tracking Research and Applications (ETRA '26), June 01–04, 2026, Marrakesh, Morocco.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3797246.3806216

1 Introduction and State of the Art

Eye tracking measures gaze direction and eye movements, providing insight into visual attention and human–computer interaction. It has become a key tool in fields including psychology, neuroscience, HCI, market research, and medical diagnostics [Duchowski 2017; Mele and Federici 2012; Wade and Tatler 2005].

Modern eye trackers often rely on video-based approaches, where cameras image the eyes and extract features such as the pupil center or corneal reflections [Crane and Steele 1985; Dodge and Cline 1901]. Although accurate and widely adopted, these systems require substantial computational resources, continuous image acquisition, and relatively bulky optics. Their high power consumption, sensitivity to environmental conditions, and mechanical footprint pose significant challenges for integration into lightweight wearable smart glasses, where constraints on size, battery life, and aesthetics are critical.

To enable pervasive and unobtrusive eye tracking, low-power camera-free alternatives have gained increasing interest. Among them, Photosensor Oculography (PSOG) relies on infrared LEDs and photodiodes to measure reflected light from the eye [Kumar and Krol 1992; Van Den Berg and Spekreijse 1997]. Because the reflected infrared signal varies systematically with eye orientation, PSOG can map these variations to gaze direction with minimal computational effort and very low power consumption, while also supporting high sampling rates and low latency suitable for wearable and embedded platforms.

More broadly, a wide range of eye-tracking technologies exists. Video-based approaches remain dominant, including feature-based systems such as PCCR [Cornsweet and Crane 1973; Crane and Steele 1985] and appearance-based machine-learning methods [Larrazabal et al. 2019; Zhang et al. 2015], but their processing demands and power consumption limit integration into consumer smart eyewear. Other non-imaging alternatives, such as electro-oculography [Estrany and Fuster-Parra 2022] and scleral search coils [Sprenger et al. 2008], avoid image processing but introduce limitations in comfort, drift, or invasiveness. Dynamic Vision Sensors have also shown promise for high-speed gaze estimation [Angelopoulos et al. 2021], although their processing pipeline remains complex for embedded smart glasses.

Within this landscape, PSOG stands out for its simplicity, scalability, and extremely low power requirements. Its challenges relate primarily to sensor placement, sensitivity to frame slippage, inter-user anatomical differences, and the nonlinear relationship between reflectance patterns and gaze direction [Palmero et al. 2023; Pezzoli et al. 2026; Van Den Berg and Spekreijse 1997]. Even small sensor displacements relative to the eye can significantly alter reflectance patterns and degrade calibration accuracy, as shown in parametric analyses of PSOG designs [Rigas et al. 2018]. Conventional approaches often rely on regression models to estimate continuous gaze angles, but recent work highlights the effectiveness of classification-based strategies [Pettenella et al. 2025; Pezzoli et al. 2025], particularly lightweight distance-based classifiers for zone-based gaze estimation [Pettenella et al. 2025]. In this perspective, discrete gaze-zone estimation is not intended to replace continuous eye tracking, but rather to provide a low-power and embedded-friendly solution for applications in which coarse spatial gaze information is sufficient, including context-aware interaction, menu navigation, attention-aware interfaces, and region-of-interest detection in pervasive smart-eyewear settings.

Building on these advantages, this article investigates distance-based classifiers for discrete gaze-zone detection using an embedded PSOG prototype. The work combines simulation-driven hardware design, optimized illumination geometry, and tailored feature extraction to enable reliable signal acquisition in a fully integrated smart-glasses form factor. Within this framework, lightweight calibration-driven classifiers based on Euclidean, standardized Euclidean, and Mahalanobis distances for embedded deployment are compared. Experiments on synthetic eyes, robotic systems, and human subjects provide a multi-stage validation of the approach and highlight the trade-off between classification robustness and computational cost in battery-powered wearable eye tracking.

2 Embedded PSOG Prototype

The proposed eye-tracking system is implemented on a battery-powered smart-eyewear prototype designed for continuous operation under the constraints of wearable devices. The Photosensor Oculography (PSOG) front-end integrates four infrared LEDs and four photodiodes per lens, embedded within the frame and positioned through simulation to maximize sensitivity to eye movements while remaining unobtrusive [Pezzoli et al. 2026]. The PSOG working principle is illustrated in Figure 1.

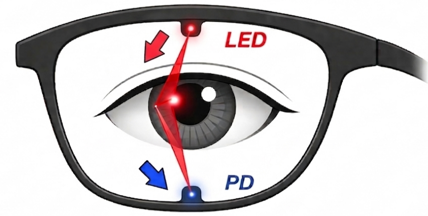

Figure 1: PSOG operating principle: The LED emits infrared (IR) light that reflects off the eye and is captured by the photodiode (PD). For clarity, only one LED and one PD are illustrated.

Infrared reflectance is acquired using a sequential illumination strategy coordinated by the microcontroller. Each LED is activated in turn while the four photodiodes are sampled, producing sixteen measurements per lens. An additional dark acquisition with all LEDs off is used to compensate for ambient illumination and electronic offsets. After dark subtraction, measurements from both lenses form a 32-dimensional feature vector per gaze sample. The acquisition pipeline is highly efficient, generating one feature vector every millisecond and supporting inference rates of approximately 1 kHz [Pezzoli et al. 2025].

Preprocessing is limited to lightweight temporal smoothing, averaging the last five samples of each feature to improve noise robustness without adding computational cost. The resulting representation preserves the infrared reflectance spatial structure needed for reliable gaze-zone classification while maintaining low latency.

The prototype also integrates auxiliary sensors, including a time-of-flight sensor, a proximity sensor, an IMU, and a magnetometer. These sensors are not used directly in the classification methods presented in this work but could enable future multimodal extensions.

All preprocessing, feature extraction, and classification run entirely on-board under the ThreadX real-time operating system, with Bluetooth available for data and result streaming. The system operates for about 15 hours on a 200 mAh battery, demonstrating the practicality of PSOG-based gaze tracking on embedded wearable platforms.

3 Distance-Based Gaze Classification

This work adopts calibration-driven distance-based classifiers to map 32-dimensional PSOG feature vectors to discrete gaze zones. Each class is represented by simple statistics computed during a short per-session calibration, and new samples are assigned to the closest class in feature space. Compared with learned models, this formulation enables rapid per-user calibration and negligible inference cost, making it well-suited to always-on embedded eyewear. While lightweight learning-based approaches could in principle be applied to PSOG features, they generally require an explicit training or adaptation stage whose computational and memory demands are less compatible with on-device personalization in

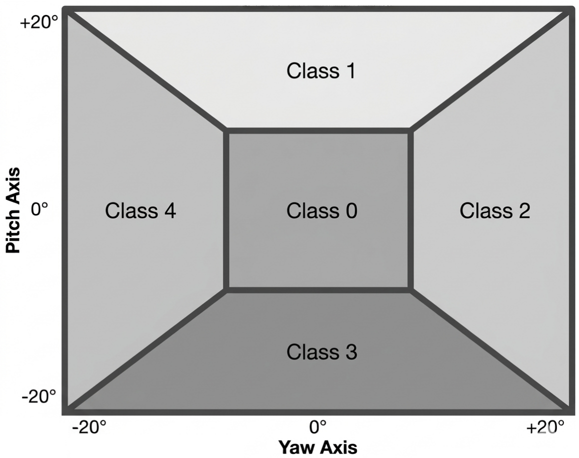

Figure 2: The division in five classes of the visual field.

resource-constrained smart eyewear. Lightweight adaptive refinement remains a possible extension for ambiguous regions of the feature space.

At the beginning of each session, the user performs a brief calibration by fixating a predefined set of targets on a display, one for each class/gaze zone. In the configuration considered here, five poses are used (up, down, left, right, center), although the same framework can be extended to denser grids. During each fixation, the embedded prototype streams the features at high rate and for every time step, a 32-dimensional vector is produced (16 corrected measurements per eye as described in Section 2). For each class c, this yields a cloud of feature vectors $\{x_c^{(n)}\}_{n=1}^{N_c}$ that characterizes the PSOG response associated with that gaze zone. From these calibration samples, class statistics are computed and stored on-board. A schematic of how the viasual field is divided into five gaze-zones is depicted in Figure 2.

Within this framework, three distance measures were evaluated: Euclidean distance, standardized Euclidean distance, and Mahalanobis distance [Duda et al. 2000]. All of them operate on the same 32-D feature vectors but rely on different calibration statistics and induce different decision boundaries.

The discrete formulation adopted in this work is a deliberate trade-off between expressiveness and embedded feasibility. By representing gaze as membership in a limited set of spatial zones, the approach enables low-complexity inference, short calibration, and robust real-time operation on battery-powered smart eyewear, while remaining suitable for the application scenarios outlined in Section 1. At the same time, it does not aim to replace continuous high-precision eye tracking in applications such as reading analysis or precise foveated rendering. Future extensions could increase the density of gaze zones or combine the present classification framework with local regression or hierarchical refinement strategies to support finer-grained gaze estimation.

3.1 Euclidean distance

The simplest classifier treats each class c as a centroid in feature space. During calibration, the mean vector $\mu_c \in \mathbb{R}^{32}$ is computed as:

$$\mu_c = \frac{1}{N_c} \sum_{n=1}^{N_c} x_c^{(n)} \tag{1}$$

Given a test vector x, its Euclidean distance to class c is:

$$d_c^{\mathrm{EUC}} = \|x - \mu_c\|_2 = \sqrt{\sum_{i=1}^{32} (x_i - \mu_{c,i})^2}. \tag{2}$$

and classification is performed by nearest centroid:

$$\hat{c}(x) = \arg\min_c d_c^{\mathrm{EUC}}. \tag{3}$$

This approach is computationally trivial and requires only one mean vector per class. However, it implicitly assumes that all feature dimensions have comparable variance and ignores correlations between channels, which may be suboptimal when some LED–PD pairs are noisier or more redundant than others [Duchowski 2017].

3.2 Standardized Euclidean distance

To mitigate the effect of unequal feature variances, a standardized Euclidean metric is also considered. In this case, calibration provides a single variance estimate σ_i^2 for each feature dimension i, typically computed by pooling all calibration vectors across classes:

$$\sigma_i^2 = \mathrm{Var}\big(\{x_i^{(n)}\}_n\big). \tag{4}$$

The distance between a test vector x and class centroid μ_c is then defined as:

$$d_c^{\mathrm{SEUC}} = \sqrt{\sum_{i=1}^{32} \frac{(x_i - \mu_{c,i})^2}{\sigma_i^2}}. \tag{5}$$

Dimensions with larger variability are down-weighted, leading to a more balanced contribution of all channels to the decision rule. The classifier is the same as in the Euclidean case, expressed by equation 3. This metric still relies only on per-feature variances and does not require covariance matrices, keeping both calibration and inference lightweight.

3.3 Mahalanobis distance

The most expressive metric evaluated is the Mahalanobis distance, which explicitly incorporates the covariance structure of each class. For a class c, calibration estimates both a mean vector μ_c and a covariance matrix $\Sigma_c \in \mathbb{R}^{32\times32}$:

$$\mu_c = \frac{1}{N_c} \sum_{n=1}^{N_c} x_c^{(n)}, \qquad \Sigma_c = \frac{1}{N_c - 1} \sum_{n=1}^{N_c} \big(x_c^{(n)} - \mu_c\big)\big(x_c^{(n)} - \mu_c\big)^\top, \tag{6}$$

The squared Mahalanobis distance between a test vector x and class c is:

$$\big(d_c^{\mathrm{MAH}}\big)^2 = (x - \mu_c)^\top \Sigma_c^{-1} (x - \mu_c). \tag{7}$$

Intuitively, this metric measures how many "standard deviations" x lies away from the class centroid when distances are computed in a whitened space induced by Σ_c. Unlike Euclidean metrics, Mahalanobis distance automatically down-weights noisy directions and accounts for cross-feature correlations, which is particularly important in the binocular PSOG setting where multiple LED–PD channels may be partially redundant or highly correlated [Duda et al. 2000]. In summary, the Euclidean and standardized Euclidean classifiers require only class centroids and per-feature variances, leading to extremely simple calibration and inference routines. The Mahalanobis classifier extends this framework by also modeling class covariances, achieving higher robustness to correlated noise and inter-channel redundancy at the cost of slightly increased, yet still embedded-friendly, complexity.

4 Experimental Protocol on Robotic Head

Preliminary evaluation of the proposed distance-based classification strategies was conducted using a controlled and repeatable protocol designed to characterize performance under realistic PSOG operating conditions while minimizing biological variability. Experiments were carried out on the *Ami* humanoid robotic head, whose mechanical and sensing characteristics are detailed in [Pezzoli et al. 2025]. Ami was selected not only for its stable and human-like ocular geometry, but also because it has been previously employed for eye-tracking validation in a few previous works [Kushina et al. 2025; Pezzoli et al. 2025]. The robotic platform eliminates confounding factors such as blinking, micro-saccades, and involuntary head motion, making it well-suited for comparing classification strategies before human validation.

During each session, Ami was commanded to fixate five discrete gaze zones: *up, down, left, right,* and *center*. A brief calibration phase was first performed, during which each target was fixated for five seconds. Calibration data were used to compute class statistics required by the distance-based classifiers, including class means and, when applicable, covariance matrices. All statistics were derived exclusively from on-board PSOG signals, and computed on the prototype.

Following calibration, randomized gaze trials were executed with the five targets presented in shuffled order. Each fixation was held for a fixed duration while feature vectors and classification outputs were continuously generated. Ground-truth labels were obtained directly from the commanded eye rotations. No recalibration was performed during testing, ensuring that the evaluation reflected robustness to noise, illumination changes, and mechanical deviations.

To assess sensitivity to eyewear placement, the protocol was repeated across three datasets corresponding to distinct eyewear mounting configurations on the robotic head. These perturbations emulate re-wearing effects in which the glasses are removed and worn again with slightly different positioning relative to the eyes. Since PSOG is known to be sensitive to millimeter-scale displacements [Rigas et al. 2018], these repeated configurations provide an initial assessment of robustness to moderate cross-session sensor displacement. Although they do not capture dynamic frame slippage during natural head motion, they capture realistic variability introduced by re-wearing the device.

The robotic-head experiments were used to benchmark alternative distance metrics, feature normalization strategies, and inference configurations over the full reachable ocular range ($\pm 25^\circ \times \pm 20^\circ$). The most effective approaches identified in this phase were subsequently evaluated on human subjects.

5 Human Dataset and Validation Experiments

Following certification of the prototype, a dedicated human-subject data acquisition campaign was conducted to validate the selected approaches under realistic operating conditions and to assess robustness to illumination variability, eye asymmetries, and natural eye motion. Eye-tracking data were collected using the PSOG-based prototype, with an EyeLink 1000 Plus (SR Research, Ontario) eye-tracking system as the ground-truth reference.

Stimulus presentation and event logging were controlled using PsychoPy. Participants were seated in front of a monitor, with head motion minimized using a chin rest to ensure a stable viewing geometry during acquisition. Consequently, dynamic frame slippage was not explicitly evaluated in the human experiments and is only partially explored through the repeated mounting configurations of the robotic protocol.

Data acquisition was performed under three ambient visible-light conditions, namely normal indoor lighting, bright lighting intended to simulate strong ambient illumination, and a reduced visible-light condition representative of low-light usage scenarios. During the normal-light condition, participants underwent a 5-points calibration procedure like the one performed on the robotic head, together with the EyeLink calibration, to establish accurate gaze mapping. For each illumination condition, two experimental tasks were performed: a fixation task on a 5×5 grid of visual targets, used to acquire static gaze samples, and a smooth pursuit task in which participants followed a continuously moving target to capture dynamic gaze trajectories. Experiments were later repeated for a 9-points calibration procedure, while keeping the same experimental tasks, in order to assess the impact of calibration density on gaze estimation accuracy. The final dataset features 29 subjects (7 females) with different skin complexions and ages from 21 to 37 years old.

During all acquisitions, the EyeLink reference tracker remained active at its minimum illumination setting, thereby introducing infrared illumination under all tested conditions. Consequently, the reduced visible-light condition should not be interpreted as IR-free. In the feature extraction pipeline, an additional dark acquisition with all LEDs switched off is subtracted from the active measurements, mitigating constant IR ambient and background components, including possible IR offsets from the EyeLink. While this compensation reduces static offsets, it may not fully eliminate the effects of external IR illumination, particularly if they alter the photodiode operating range or the reflectance baseline.

5.1 Data processing

Raw PSOG signals were processed to extract the same 32-dimensional feature vectors used in the robotic experiments, corresponding to all PD–LED combinations from both eyes. To isolate valid gaze segments and suppress transient artifacts, blinks and saccades were removed using a motion-energy-based filtering approach. The temporal derivative of each feature signal was computed, and the L2 norm of the resulting velocity vector was used as a continuous measure of eye motion intensity. A participant-specific detection threshold was estimated from an initial 5-second steady fixation period and set to $\mu + 3\sigma$ of the baseline energy. Samples exceeding this threshold were classified as saccades or blinks and removed, with additional temporal padding applied to eliminate onset and offset artifacts. The resulting cleaned signals retained only stable

Table 1: Results on Ami (five-class classification over the entire $\pm 25^\circ \times \pm 20^\circ$ gaze range).

Method	Dataset 1	Dataset 2	Dataset 3	Avg.
KNN	0.79	0.64	0.66	0.70
GMM	0.78	0.64	0.77	0.73
Euclidean	0.75	0.64	0.68	0.69
Std. Euclidean	0.64	0.54	0.64	0.61
Mahalanobis	**0.81**	**0.68**	**0.69**	**0.73**

Table 2: Human-subject mean classification accuracy for Mahalanobis classifier under different illumination conditions.

Illumination	5 classes	9 classes
Bright Lighting	0.62	0.59
Normal Indoor Lighting	**0.69**	**0.72**
Reduced Visible-Light	0.60	0.54

fixation and smooth pursuit segments suitable for training and evaluation.

To evaluate the impact of asymmetric eye behavior and sensor-level differences between eyes, two inference configurations were tested. In the unified configuration, a single classifier was trained using concatenated features from both eyes. In the split-eye configuration, two independent classifiers were trained on left and right eye features, respectively, and their outputs were fused by selecting the result with the highest confidence score, defined as

$$\text{score} = 1 - \frac{d_{\text{closest}}}{d_{\text{second closest}}}. \tag{8}$$

This formulation was designed to increase robustness to sensor misalignment, partial occlusions, and anatomical asymmetries between eyes, while reducing computational complexity by operating on lower-dimensional inputs.

Dimensionality reduction was investigated exclusively in the human-subject experiments to balance accuracy, robustness, and embedded feasibility. Three strategies were explored: Principal Component Analysis, variance maximization (selects the features with the highest individual variability), and inter-class difference maximization (selects features that maximize the ratio of between-class to within-class variance). These methods were tested at target dimensionalities of 4, 8, 12, 16, 24, and 32 features. Different subsets of calibration data corresponding to specific spatial patterns were considered, including local neighborhood and radial configurations, to assess their effect on discriminability and generalization. For brevity, only the most representative and performing configurations are reported in Section 6.

6 Results

Distance-based gaze classification on the Ami humanoid robot was first evaluated to validate the choice of metric in a controlled five-class setting. The Mahalanobis distance classifier was compared against Euclidean and standardized Euclidean distances, as well as classical baselines (KNN and GMM).

6.1 Results on Ami Robot

Table 1 reports classification accuracies across three independently acquired datasets corresponding to different eyewear mounting conditions. Mahalanobis distance achieved the highest or near-highest accuracy, outperforming alternative distance metrics and classical baselines. These results confirm the benefit of explicitly modeling feature correlations when gaze-related class distributions become elongated and anisotropic in the high-dimensional PSOG feature space.

6.2 Results on Human Subjects

Based on the robot experiments, Mahalanobis distance was adopted for large-scale evaluation on human subjects. Table 2 reports average classification accuracy across illumination conditions for five- and nine-class gaze layouts. Under normal lighting, accuracy reached 0.67–0.70 for the five-class task and 0.71–0.73 for the nine-class task. The slightly higher performance of the nine-class configuration is attributed to its spatial layout, which reduces ambiguity near the central region. A recurrent source of error is the ambiguity between adjacent gaze zones when the true gaze falls near a class boundary. In these cases, the PSOG feature vectors may lie close to multiple class centroids, leading to occasional confusion between neighboring regions. This behavior is consistent with the discrete formulation of the problem and is less critical in applications such as menu navigation or context-aware interaction, where coarse spatial selection is often sufficient, but it would be more limiting in tasks requiring fine-grained or continuous gaze estimation.

Across illumination conditions, split-eye inference provided a consistent improvement of about 2–3% in accuracy, suggesting that independent modeling of left and right eye signals helps mitigate asymmetries due to alignment, occlusions, and eye-specific illumination differences. Performance differences across conditions are likely influenced by pupil dynamics, illumination-induced signal variations, and residual interactions with the acquisition setup. In particular, the reduced visible-light condition should be interpreted cautiously, since the active IR reference tracker means that it does not correspond to a fully IR-dark environment. Nevertheless, correct gaze region identification was retained in the majority of cases.

Dimensionality reduction was then explored on human data to improve robustness and computational efficiency. Figure 3 shows subject-wise accuracy for the nine-class task under normal lighting. While median performance is high, a few low-performing participants substantially affect the mean, indicating marked inter-subject variability. PCA with 24 retained dimensions combined with Mahalanobis distance achieved stable performance across illumination conditions (Table 3, first row) at a cost of 6168 MAC operations. A lighter configuration based on inter-class difference maximization with 14 selected features and Euclidean distance achieved competitive accuracy at only 140 MAC operations (Table 3, last row), making it attractive for low-power embedded deployment.

Overall, reduced feature representations preserved classification performance while substantially lowering computational cost, supporting flexible accuracy–power trade-offs across hardware platforms.

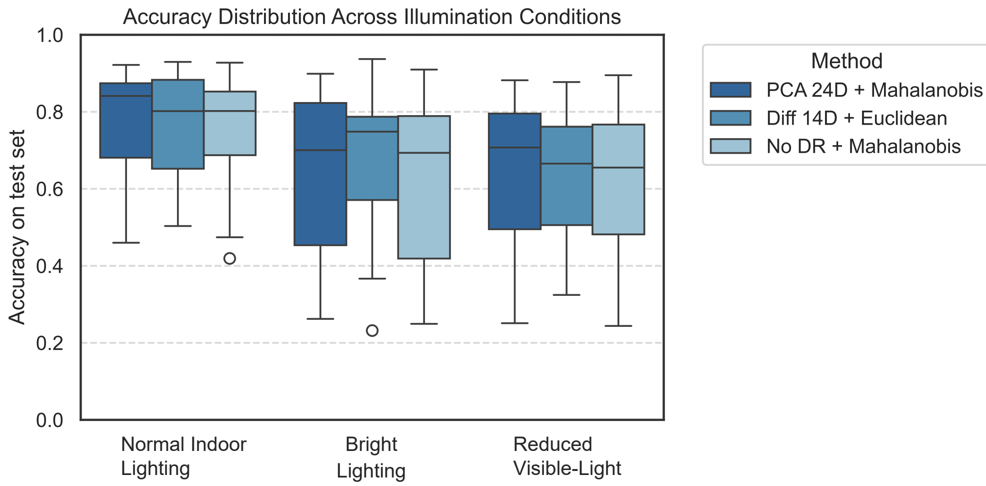

Figure 3: Subject-wise accuracy distribution for selected best methods for nine-class gaze classification under different lighting conditions.

Table 3: Performance of selected reduced-dimensionality configurations across illumination conditions.

Method	MACs	Bright Lighting	Normal Indoor Lighting	Reduced Visible-Light
PCA (24D) + Mahalanobis	6168	0.65	**0.80**	0.61
Diff. (14D) + Euclidean	140	0.68	**0.79**	0.61

7 Conclusions

This work demonstrates that Photosensor Oculography (PSOG) combined with lightweight distance-based classifiers enables real-time gaze-zone classification on embedded smart eyewear. Experiments on a controlled robotic platform identified Mahalanobis distance as the most robust metric, and human-subject validation confirmed similar trends across five- and nine-zone layouts and varying illumination conditions. Split-eye modeling improved robustness, while dimensionality reduction reduced computational cost with limited impact on accuracy.

The complete Mahalanobis calibration and inference pipeline was implemented on an STM32U5B9 microcontroller, where inference runs in under 1 ms with a minimal memory footprint, supporting continuous wearable operation. Overall, these results establish Mahalanobis-based distance classification as an efficient and embedded-friendly approach for PSOG-based gaze estimation. At the same time, a known challenge of PSOG systems is their sensitivity to sensor displacement relative to the eye: while the robotic experiments partially addressed re-wearing variability through repeated mounting configurations, dynamic frame slippage during natural head motion was not explicitly evaluated and remains an important direction for future work.

8 Privacy and ethics statement

The methods and findings in this study are based on data collected in controlled laboratory settings using a robotic system and human participants. All human-subject procedures were approved by the appropriate ethics committee before data collection and conducted in accordance with institutional and ethical guidelines. Participants provided informed consent, and no personal or sensitive data beyond eye-movement recordings were collected. The study poses no significant risks to privacy, fairness, safety, or human rights beyond those addressed through ethical review. We foresee minimal societal risks and believe the work makes a positive contribution to eye-tracking research.

References

Anastasios N. Angelopoulos, Julien N.P. Martel, Amit P. Kohli, Jörg Conradt, and Gordon Wetzstein. 2021. Event-Based Near-Eye Gaze Tracking Beyond 10,000 Hz. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (2021), 2577–2586. doi:10.1109/TVCG.2021.3067784

Tom N Cornsweet and Hewitt D Crane. 1973. Accurate two-dimensional eye tracker using first and fourth Purkinje images. *Journal of the Optical Society of America* 63, 8 (1973), 921–928.

Hewitt D Crane and Carroll M Steele. 1985. Generation-V dual-Purkinje-image eye-tracker. *Applied optics* 24, 4 (1985), 527–537.

Raymond Dodge and Thomas Sparks Cline. 1901. The angle velocity of eye movements. *Psychological Review* 8, 2 (1901), 145.

Andrew T. Duchowski. 2017. *Eye Tracking Methodology: Theory and Practice* (3rd ed.). Springer, Cham, Switzerland. doi:10.1007/978-3-319-57883-5

Richard O. Duda, Peter E. Hart, and David G. Stork. 2000. *Pattern Classification (2nd Edition)* (2 ed.). Wiley-Interscience.

Bartomeu Estrany and Pilar Fuster-Parra. 2022. Human Eye Tracking Through Electro-Oculography (EOG): A Review. In *International Conference on Cooperative Design, Visualization and Engineering*. Springer, 75–85.

Arvind Kumar and Glenn Krol. 1992. Binocular infrared oculography. *The Laryngoscope* 102, 4 (1992), 367–378.

Nadezhda Kushina, Ko Watanabe, Aarthi Kannan, Ashita Ashok, Andreas Dengel, and Karsten Berns. 2025. SensHRPS: Sensing Comfortable Human-Robot Proxemics and Personal Space With Eye-Tracking. arXiv:2512.08518 [cs.RO] https://arxiv.org/abs/2512.08518

Agostina J Larrazabal, CE García Cena, and César Ernesto Martínez. 2019. Video-oculography eye tracking towards clinical applications: A review. *Computers in biology and medicine* 108 (2019), 57–66.

Maria Laura Mele and Stefano Federici. 2012. Gaze and eye-tracking solutions for psychological research. *Cognitive Processing* 13, 1 (01 Aug 2012), 261–265. doi:10.1007/s10339-012-0499-z

Cristina Palmero, Oleg V Komogortsev, Sergio Escalera, and Sachin S Talathi. 2023. Multi-rate sensor fusion for unconstrained near-eye gaze estimation. In *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications*. ACM, 1–8.

Alberto Pettenella, Daniele Maria Crafa, Jacopo Spagnoli, Carlo Pezzoli, Marco Paracchini, Susanna Di Giacomo, Carlo Fiorini, Sean Byrne, Tommaso Ongarello, Luca Merigo, and Marco Carminati. 2025. Development of a Low-Power Wearable Eye Tracker based on Hidden Photodetectors. In *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications (ETRA '25)*. Association for Computing Machinery, New York, NY, USA, Article 117, 7 pages. doi:10.1145/3715669.3727347

Carlo Pezzoli, Marco Brando Mario Paracchini, Daniele Maria Crafa, Marco Carminati, Luca Merigo, Tommaso Ongarello, and Marco Marcon. 2026. Design and Implementation of an Eyewear-Integrated Infrared Eye-Tracking System. *Sensors* 26, 7 (2026), 2065. doi:10.3390/s26072065

Carlo Pezzoli, Emanuele Santoro, Marco Brando Mario Paracchini, Giulio Marano, Daniele Bani, Luca Francesco Raduzzi, Daniele Maria Crafa, Marco Carminati, Luca Merigo, Tommaso Ongarello, Marco Marcon, and Stefano Tubaro. 2025. Low-Power Hierarchical Network: Pervasive Eye-Tracking on Smart Eyewear. In *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications (ETRA '25)*. Association for Computing Machinery, New York, NY, USA, Article 121, 7 pages. doi:10.1145/3715669.3727346

Ioannis Rigas, Hayes Raffle, and Oleg V. Komogortsev. 2018. Photosensor Oculography: Survey and Parametric Analysis of Designs Using Model-Based Simulation. *IEEE Transactions on Human-Machine Systems* 48, 6 (2018), 670–681. doi:10.1109/THMS.2018.2807244

Andreas Sprenger, Birte Neppert, Sabine Köster, Steffen Gais, Detlef Kömpf, Christoph Helmchen, and Hubert Kimmig. 2008. Long-term eye movement recordings with a scleral search coil-eyelid protection device allows new applications. *Journal of neuroscience methods* 170, 2 (2008), 305–309.

Thomas JTP Van Den Berg and Henk Spekreijse. 1997. Near infrared light absorption in the human eye media. *Vision research* 37, 2 (1997), 249–253.

Nicholas Wade and Benjamin Tatler. 2005. *The Moving Tablet of the Eye: The origins of modern eye movement research*. Oxford University Press. doi:10.1093/acprof:oso/9780198566175.001.0001

Xucong Zhang, Yusuke Sugano, Mario Fritz, and Andreas Bulling. 2015. Appearance-based gaze estimation in the wild. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 4511–4520. doi:10.1109/CVPR.2015.7299081