International Journal of Basic and Applied Science 14 (4) (2026) 137-151

Published by: IOCSCIENCE

International Journal of Basic and Applied Science

Journal homepage: www.ijobas.pelnus.ac.id

Electrooculography Based Control of a Robotic Manipulator with Dual Cameras for Object Retrieval

Muhammad Ilhamdi Rusydi[1], Andre Paskah Gultom[2], Adam Jordan[3], Rahmad Novan Nurhadi[4], Darwison[5]

[1,2,3,4,5]Department of Electrical Engineering, Universitas Andalas, Indonesia

Article Info

Article history:

Received Oct 10, 2025
Revised Nov 07, 2025
Accepted Mar 08, 2026

Keywords:

Electrooculography;
Robot Manipulator;
Image Processing;
Human-Computer Interface;
Dual-Camera Integration.

ABSTRACT

This study presents an assistive control system for a four-degree-of-freedom (4-DoF) robotic manipulator that integrates image-based spatial perception with electrooculography (EOG)-based human–machine interaction for three-dimensional object retrieval. The system is motivated by the need for intuitive, non-contact assistive technologies to support individuals with severe motor impairments, such as tetraplegia, in performing basic manipulation tasks. The proposed framework employs an orthogonal dual-camera vision configuration to achieve explicit 3D target localization, where planar object positions on the XY plane and depth along the Z axis are estimated using focal length–based geometric modeling. User commands are generated through an EOG interface, in which eye movements and voluntary blinks are classified using a K-Nearest Neighbor (KNN) algorithm to control manipulator motion. Compared to conventional assistive robotic systems that rely on depth sensors or high-degree-of-freedom manipulators, the proposed approach utilizes asymmetric monocular viewpoints and a minimal 4-DoF architecture to reduce system complexity. Experimental results demonstrate high performance, achieving average localization accuracies of 99.52% on the XY plane and 95.88% along the Z axis, as well as an EOG classification accuracy of 94.38%. Manipulation experiments confirmed reliable operation with a 100% task success rate, while task completion time and positional error increased gradually with target distance. These findings validate the feasibility of the proposed system as a low-complexity, high-accuracy assistive robotic solution for rehabilitation and human–machine interaction applications.

Corresponding Author:

Muhammad Ilhamdi Rusydi,
Department of Electrical Engineering,
Universitas Andalas,
Jl. Dr. Mohammad Hatta, Limau Manis, Kec. Pauh, Kota Padang, Sumatera Barat, 25175, Indonesia
Email: rusydi@eng.unand.ac.id

1. INTRODUCTION

The number of people with severe motor disabilities continues to increase due to chronic diseases, accidents, and population aging [1]. Tetraplegia, characterized by paralysis from the neck down, severely restricts the ability to perform basic daily activities independently [1], [2] highlighting the need for assistive technologies that can restore user autonomy.

Robotic manipulators have been widely explored for assistive applications. While robotic arms are well established in industrial settings [3], [4], [5] their adaptation to assistive and healthcare scenarios has accelerated with advances in sensing, computation, and human–machine interfaces [6],

[7]. Conventional control interfaces such as joysticks and switches [8] have been complemented by bio-signal-based approaches, including EMG [9] and EEG [10]. However, EEG-based systems are highly sensitive to noise and require complex user-specific calibration, while EMG-based systems depend on residual muscle activity and are unsuitable for individuals with severe motor impairments. Despite these advances, robust grasping and pick-and-place in unstructured environments remains challenging; reported success rates for unstructured grasping often range from about 70% to 85%, implying non-negligible failure rates during perception-to-action execution [3], [6], [11]. This limitation is particularly pronounced when object localization and grasping must accommodate varying object heights, motivating explicit three-dimensional position estimation in practical assistive settings [12].

Among bio-signal interfaces, Electrooculography (EOG) has emerged as a promising modality for assistive robotic control, as natural eye movements often remain accessible even in advanced motor disabilities. As summarized in Table 1, EEG-, EMG-, and EOG-based interfaces differ in key usability trade-offs, including latency, cognitive load, calibration time, and robustness. EEG-based systems typically impose high cognitive load and require extensive user-specific calibration [6], [13], while EMG-based approaches depend on voluntary muscle activation and are therefore unsuitable for severe motor impairments [13]. In contrast, EOG-based interfaces offer lower cognitive load, shorter calibration time, and higher robustness, enabling intuitive command generation with minimal training and motivating their selection as a practical, user-centered control modality for assistive robotic manipulation in unstructured environments [14], [15].

Table 1. Comparative Characteristics of EEG, EMG, and EOG for Assistive Control Interfaces

Aspect	EEG	EMG	EOG
Control Source	Cortical neuronal activity	Residual muscle activity	Eye movement (ocular dipole)
Signal Amplitude	10–100 μV [11]	0.1–5 mV [9]	50–3500 μV [14]
Latency	High (hundreds of ms due to signal processing and mental decoding) [3]	Low–Moderate (tens to hundreds of ms) [9]	Low (tens of ms, rapid eye movement detection) [16]
Cognitive Load	High (mental imagery and sustained attention) [6]	Low–Moderate (muscle activation) [9]	Low (natural eye movements) [11]
Calibration Time	Long, user-specific	Moderate	Short
Robustness	Low, sensitive to noise and artifacts	Moderate affected by muscle fatigue and electrode shift	High, relatively stable and robust
Suitability for Tetraplegia	Limited (complex setup, sensitivity issues)	Not suitable (requires voluntary muscle activity)	Suitable

Table 2. Summary of previous research

Topic	Workspace Dimention	Dataset	Metric	Real-time	Auto grasp	Height variation	Limitation
Joystick control [8]	3D	N/S	Qualitative	Yes	No	No	Needs hand control
EMG control [9]	3D	User trials	Cmd accuracy + qual.	Yes	No	No	Needs residual muscle
EEG BCI control [17]	Virtual 3D	Offline + trials	BCI cmd accuracy	Limited	No	No	Artifact-prone; long calib.
EOG 3-DoF arm [18]	2D–3D (limited)	N/R	Direction recog.	Yes	No	Limited	No full pick-place eval.
EOG + dual cams [14]	3D	Limited	Qual. coverage	Yes	No	No	Restricted range

Several studies have demonstrated the feasibility of EOG-based robotic manipulator control. Early studies focused on 3-DoF manipulators operating in two-dimensional workspace [14], while later works extended control to three-dimensional space using dual-camera configurations [19] and integrated vision-based object detection to enable basic pick-and-place tasks. Despite these advances, most existing EOG-based systems remain limited by low degrees of freedom, restricted workspace coverage, and the lack of explicit object height handling, leading to inconsistent grasping performance when objects are located at different heights or outside the primary field of view. A comparative summary of prior studies with respect to workspace dimensionality, autonomous grasping capability, and height variation handling is provided in Table 2.

Unlike previous studies that rely on low-degree-of-freedom manipulators and limited visual coverage, this work proposes an EOG-based assistive manipulation framework employing a four-degree-of-freedom (4-DoF) robotic manipulator integrated with two orthogonally positioned cameras for explicit three-dimensional workspace modeling. The system enables height-aware object localization and teleoperated pick-and-place tasks without autonomous grasping, while maintaining system simplicity. Performance is evaluated through experimental trials using quantitative metrics, including task completion rate, positioning accuracy, and real-time response feasibility. By extending workspace coverage to vertical and rotational motion with minimal cognitive and calibration burden, the proposed approach enhances the practicality and adaptability of EOG-based assistive robotic manipulation.

2. RESEARCH METHOD

This study adopts a constraint-driven design approach for assistive robotic manipulation, prioritizing functional sufficiency under physical, cognitive, and computational limitations rather than maximal mechanical complexity [20]. A four-degree-of-freedom (4-DoF) manipulator is selected as the minimal configuration capable of supporting three-dimensional pick-and-place tasks—base rotation, planar reach, and vertical motion—while preserving robustness and usability in assistive applications.

Although higher-degree-of-freedom manipulators (5–6 DoF) provide orientation control and redundancy, they introduce increased mechanical complexity, calibration sensitivity, computational overhead, and user cognitive load that are unnecessary for the targeted tasks. Consequently, the 4-DoF architecture represents an effective trade-off between task completeness and system simplicity.

Three-dimensional target localization is performed using an orthogonal dual-camera setup, enabling focal length–based geometric estimation without dedicated depth sensors [21], thereby reducing cost and computational demand while maintaining real-time operation. User interaction is implemented through an Electrooculography (EOG)-based interface, where gaze directions encode motion commands and voluntary blinks serve as activation signals, minimizing physical and cognitive effort for users with severe motor impairments.

System performance is evaluated using task-level metrics, including positional accuracy, pick-and-place success rate, and real-time execution stability, emphasizing practical manipulation performance over signal recognition accuracy alone. As shown in Figure 1, perception, intention confirmation, and actuation are integrated into a closed-loop assistive control framework that completes the full manipulation cycle from target localization to object placement and safe return to the HOME position.

Figure 1. Four steps in developing a robotic manipulator control system using electrooculography.

2.1 Camera Calibration and Spatial Configuration

Both cameras were calibrated prior to experimentation using Zhang's planar checkerboard method to estimate intrinsic parameters and lens distortion coefficients [22]. Multiple checkerboard views were

processed via homography-based optimization to obtain accurate geometric camera models. Calibration accuracy was evaluated using the root mean square (RMS) reprojection error, defined as the Euclidean distance between detected and reprojected checkerboard corners [23]. For both cameras, the RMS reprojection error was approximately 1.2 pixels, providing a quantitative indicator of geometric consistency for monocular vision-based manipulation.

Figure 2 illustrates the camera calibration and object localization setup. The planar checkerboard serves as the calibration target, while the red square represents the object used for distance estimation. Horizontal and vertical reference lines denote the image coordinate axes employed for pixel-based measurement and spatial mapping [23].

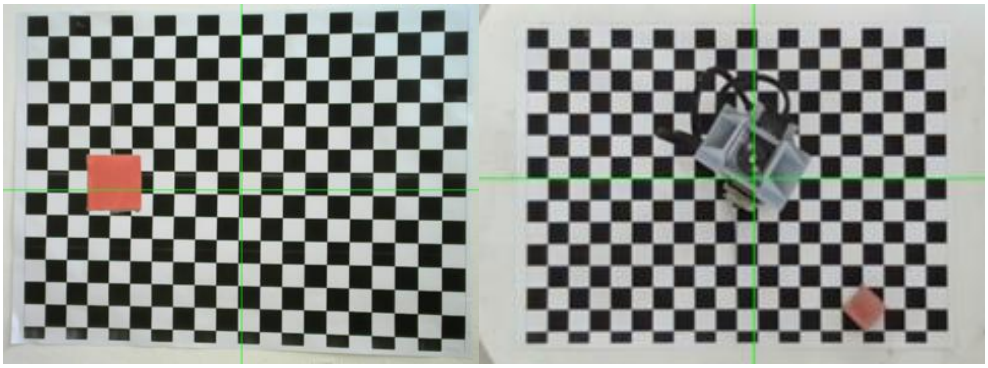
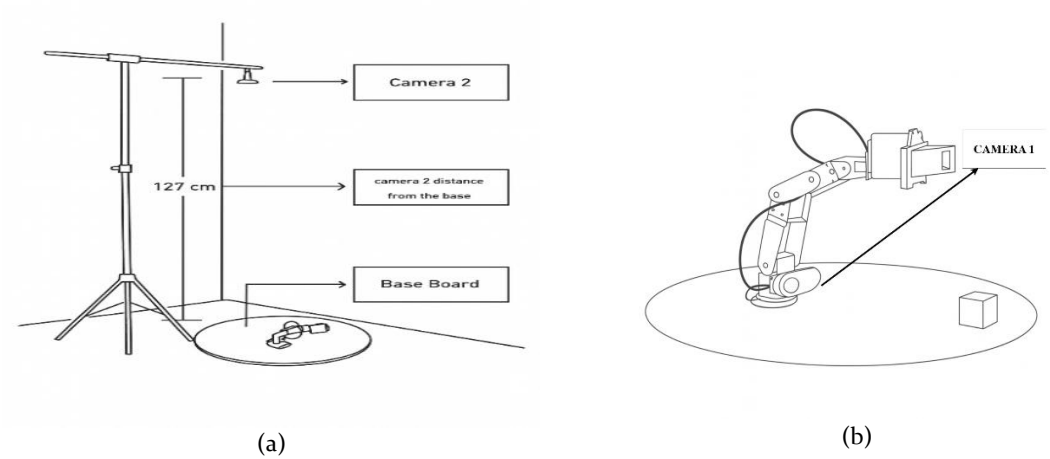


Figure 2. Camera Calibration Setup Using a Planar Checkerboard Pattern

The distance estimation model employs calibrated focal length parameters derived from the intrinsic camera matrix, rather than nominal manufacturer specifications. Prior to coordinate computation, all images were undistorted using the estimated distortion coefficients to minimize systematic geometric errors and ensure consistency in pixel-to-world transformation.

The vision system utilizes two asymmetrically placed monocular cameras, with the top-view camera estimating object distance along the Z-axis (height) and the side-view camera estimating planar X–Y coordinates. Since the cameras do not form a parallel stereo configuration with a fixed baseline, depth estimation is performed using calibrated monocular geometry from complementary viewpoints rather than stereo disparity.



(a) (b)

Figure 3. Illustration of camera placement in the system: (a) top camera, (b) bottom camera

Figure 3 shows the camera placement configuration. As illustrated in Figure 3(a), the top camera is positioned above the workspace and used for Z-axis distance estimation based on a calibrated focal

length–based model, providing reliable depth estimation with minimal occlusion. Figure 3(b) presents the side camera, which captures lateral object displacement for planar (X–Y) position estimation. By combining measurements from the top and side cameras, the system enables explicit three-dimensional (X–Y–Z) coordinate estimation while maintaining computational efficiency suitable for real-time assistive manipulation.

Object distance relative to the camera is estimated using the geometric similarity principle, based on the relationship between actual object size, projected image size, and calibrated camera parameters, as illustrated in Figure 4 and reported in previous studies [21]. The required parameters include camera resolution, sensor dimensions, and calibrated focal length. In this study, a camera with a resolution of 640 × 480 pixels, a sensor size of 3.58 mm × 2.02 mm, and a focal length of 4 mm was used.

Figure 4. Illustration taking image on camera

While nominal camera specifications are provided for reference, all distance estimation computations rely exclusively on calibrated intrinsic parameters obtained through Zhang's method [23]. Accordingly, the proposed vision system continues to employ a focal length–based model, complemented by distortion correction and a quantified RMS reprojection error, ensuring sufficient geometric accuracy for real-time assistive robotic manipulation without the added complexity of stereo disparity approaches.

2.2 Participants and Electrooculography Data Collection Protocol

This study adopts a multi-user, subject-dependent evaluation protocol involving ten healthy adult participants (aged 22–35 years) with normal or corrected-to-normal vision and no reported neurological or ophthalmological disorders. Participants were required to perform reliable voluntary gaze shifts and blink actions, while individuals exhibiting excessive involuntary blinking, unstable baseline EOG signals, or discomfort during electrode placement were excluded. All participants provided informed consent prior to data acquisition. Healthy subjects were used to assess system feasibility, robustness, and repeatability under controlled conditions, without claiming direct clinical applicability. EOG data were analyzed independently for each participant, with no cross-subject training or validation, emphasizing subject-specific performance consistency rather than population-level generalization.

Prior to recording, participants were trained to perform predefined eye-movement commands, including horizontal and vertical gaze and voluntary blinks. Recording sessions were limited to reduce fatigue effects, with rest intervals provided between trials and additional breaks allowed when performance degradation or discomfort was observed.

EOG signals were acquired using four disposable surface electrodes arranged in a standard EOG configuration (Figure 5), consisting of two active electrodes for horizontal and vertical movements, one reference electrode, and one ground electrode [7]. Electrode placement was standardized and inspected before each session to minimize displacement and impedance variation. Baseline signals were recorded at session onset, with recalibration performed when baseline drift or amplitude variation was detected.

Physiologically, EOG signals arise from the corneo-retinal standing potential, producing low-frequency components associated with eye rotation and blink activity. Voluntary gaze shifts and blinks predominantly reside below 5 Hz, whereas baseline drift occurs at very low frequencies (<1 Hz).

Accordingly, a band-pass filter of 1.06–4.97 Hz was applied to preserve physiologically relevant eye-movement components while suppressing baseline drift and high-frequency noise, consistent with established EOG studies [7], [16], [18], [24].

The recording process yielded two channels: Channel 1 (Ch1) representing horizontal eye movements and Channel 2 (Ch2) representing vertical eye movements. Table 3 summarizes the polarity patterns of EOG signals for different gaze directions and blink events.

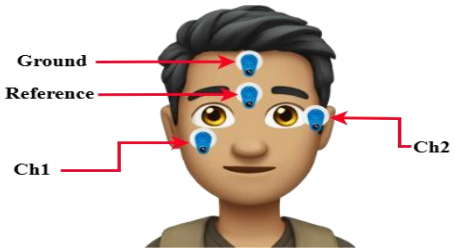


Figure 5. Electrode Placement on the User [7]

Table 3. Polarity Features Base on Type Signal Eye Movement [7]

Channel	Detection Eye Gaze				Detection Blink	
	Left	Right	Up	Down	Voluntary Blink	Involuntary Blink
Ch1	(-)	(+)	(-)	(+)	(-)	(-)
Ch2	(+)	(-)	(-)	(+)	(-)	(-)

Thresholds were determined using a deterministic, participant-specific calibration procedure performed at the beginning of each session. Baseline EOG signals were recorded for 10 s during neutral gaze to compute the mean (μ) and standard deviation (σ) for each channel. Detection thresholds were defined as $T_{upper} = \mu + 3\sigma$ and $T_{lower} = \mu - 3\sigma$, without manual tuning, ensuring consistent threshold derivation across participants and sessions.

Feature extraction focused on amplitude and period characteristics to capture gaze magnitude and temporal behavior, while threshold-crossing information enhanced robustness to baseline drift. This combined feature set enabled reliable discrimination between gaze directions and blink events, as illustrated in Figure 6.

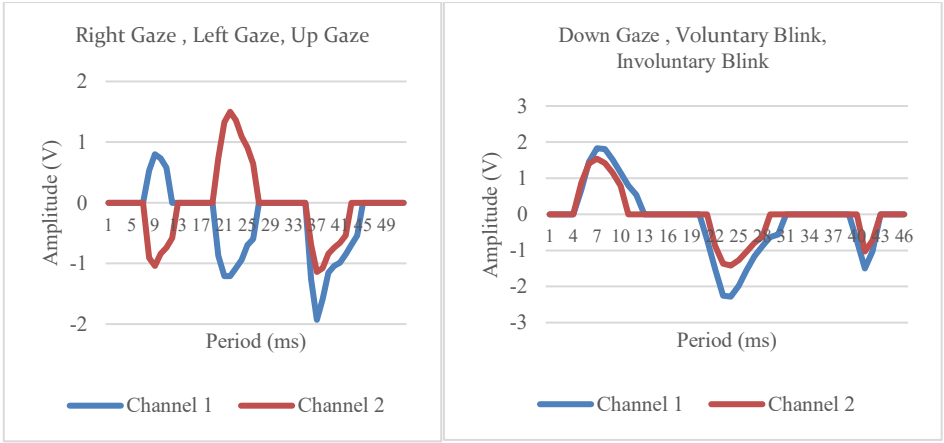


Figure 6. Signal patterns for different gaze directions and blinks

K-Nearest Neighbor (KNN) was selected as the primary classifier due to its simplicity, interpretability, and low computational cost in low-dimensional feature spaces, making it suitable for real-time assistive robotic systems. Support Vector Machine (SVM) and Random Forest (RF) were included as comparative baseline models under identical conditions, along with a rule-based classifier using polarity patterns and threshold-crossing rules as a reference [26].

An ablation study evaluated the contribution of different feature groups by comparing amplitude-only, amplitude–period, and complete feature sets using amplitude and period feature groups. Classification performance was assessed using accuracy, macro-averaged precision, recall, and F1-score. Where applicable, fold-wise macro-F1 scores were compared across feature sets using paired non-parametric tests (Wilcoxon signed-rank) to reduce sensitivity to distributional assumptions.

System-level performance was validated through controlled manipulation experiments involving object scanning, pick-and-place, and return-to-home motions. To minimize variability during proof-of-concept evaluation, end-to-end trials were conducted by one trained operator who issued EOG commands after completing the calibration session. A total of 25 trials were performed at five working distances (8–20 cm). Placement error (PE) was defined as the Euclidean distance between the final end-effector position and the intended target placement, while task success rate (TSR) was computed as the percentage of successful trials. Confusion matrix analysis was used to examine inter-class misclassification patterns in EOG-based command interpretation.

3. RESULTS AND DISCUSSIONS

3.1 Analysis of Object Distance Calculation Using a Camera

In the distance measurement process, a dual-camera system was employed to estimate object positions in three-dimensional space using a focal-length-based geometric model. Quantitative results from both cameras are summarized in Table 4, while representative visual outputs are shown in Figure 7.

For Camera 1, which was mounted horizontally to estimate object displacement along the X–Y plane (Figure 7a), the signed error ranged from –0.35 cm to 0.23 cm across 10 trials, yielding a mean accuracy of 99.52%. The mean absolute error (MAE) and root-mean-square error (RMSE) were 0.10 cm and 0.148 cm, respectively, with a standard deviation of 0.15 cm, indicating low variability across the evaluated range (Table 4). In contrast, Camera 2, mounted vertically to estimate object height along the Z-axis (Figure 7b), achieved an average accuracy of 95.88% over 10 trials. Errors ranged from 0.03–0.63 cm with a standard deviation of 0.21 cm; the MAE and RMSE were 0.28 cm and 0.343 cm, respectively, reflecting higher variability in Z-axis estimation, particularly at longer distances (Table 4).

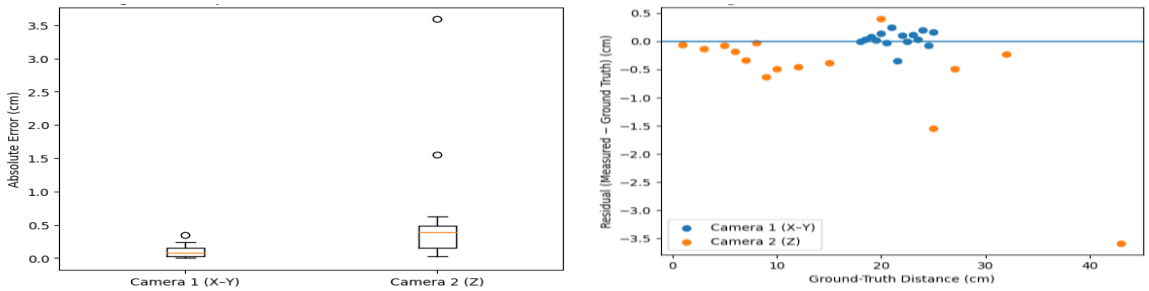
(a) (b)

Figure 7. (a)Output display camera 1, (b)Output display camera 2

To further examine the variability of distance estimation errors beyond mean accuracy values, the distribution of absolute measurement errors was analyzed. Figure 8(a) presents a boxplot of absolute errors for both Camera 1 and Camera 2.

As shown in Figure 8(a), X–Y plane measurements obtained from Camera 1 exhibit a narrow interquartile range, indicating that most errors are tightly clustered around the median value. In

contrast, Z-axis measurements obtained from Camera 2 show a wider spread and several higher-magnitude outliers, particularly at longer distances. This difference reflects the increased sensitivity of monocular depth estimation to distance variation and optical effects.



(a) (b)

Figure 8. (a) Boxplot of Distance Measurement Error, (b) Residual Plot of Distance Estimation

Table 4. Focal length measurement results

No.	Focal Length Measurement Results for X and Y Axes Using Camera 1				Focal Length Measurement Results for the Z Axis Using Camera 2			
	Target Distance (cm)	Measured Distance (cm)	Error (cm)	Accuracy (%)	Target Distance (cm)	Measured Distance (cm)	Error (cm)	Accuracy (%)
1	18	18	0	100	1	0.94	0.06	93.13
2	18.5	18.53	0.02	99.86	3	2.86	0.14	95.02
3	19	19.08	0.08	99.57	5	4.92	0.08	98.32
4	19.5	19.52	0.02	99.89	6	5.82	0.18	96.96
5	20	20.14	0.14	99.30	7	6.66	0.34	94.82
6	20.5	20.47	- 0.04	99.83	8	7.97	0.03	99.56
7	21	21.24	0.23	98.88	9	8.37	0.63	92.42
8	21.5	21.15	- 0.35	98.35	10	9.51	0.49	94.84
9	22	22.10	0.10	99.54	12	11.54	0.46	95.97
10	22.5	22.49	- 0.01	99.96	15	14.61	0.39	97.31
Average (%)				99.52	Average (%)			95.88
Standard Deviation (cm)				0.15	Standard Deviation (cm)			0.21

To analyze distance-dependent behavior in the estimation process, regression residuals were computed as the difference between measured and ground-truth distances. Residual plots for both cameras are shown in Figure 8(b). Camera 1 residuals are closely centered around zero across the evaluated distance range, indicating minimal systematic bias in X–Y estimation. In contrast, Camera 2 residuals increase in magnitude at larger distances, suggesting distance-dependent error in Z-axis estimation.

These residual patterns indicate mild non-linear behavior in the focal-length-based distance estimation model, particularly along the Z-axis. Despite camera calibration, residual lens distortion, perspective effects, and scale compression at extended distances contribute to increased error variance near the workspace boundary.

To further quantify estimation consistency, mean absolute error (MAE) and root mean square error (RMSE) were computed. For X–Y estimation using Camera 1, MAE and RMSE values of 0.10 cm and 0.148 cm indicate millimeter-level positioning error. In contrast, Z-axis estimation using Camera 2 produced higher MAE and RMSE values of 0.280 cm and 0.343 cm, reflecting the greater sensitivity of monocular depth estimation to distance variation and optical effects.

Despite these limitations, the observed error magnitudes remain acceptable for assistive robotic manipulation tasks, where moderate positional accuracy is sufficient for safe object approach

and grasping. Future work may further reduce non-linear effects through distance-dependent calibration or error compensation strategies.

3.2 Analysis of Gaze and Blink Recognition Accuracy

EOG signal classification was performed using the K-Nearest Neighbor (KNN) method. The dataset consisted of four features extracted from two EOG channels, namely Ch1 amplitude, Ch2 amplitude, Ch1 period, and Ch2 period, with each sample assigned to one of six predefined eye-movement classes. These classes correspond to right gaze, left gaze, upward gaze, downward gaze, voluntary blink, and involuntary blink.

The labeled feature data were used to train the KNN classifier, while the distribution of samples across the six classes was visualized to examine feature separability. As illustrated in Figure 9, each class is represented by a distinct color, indicating the overall distribution and overlap patterns among gaze and blink signals prior to classification.

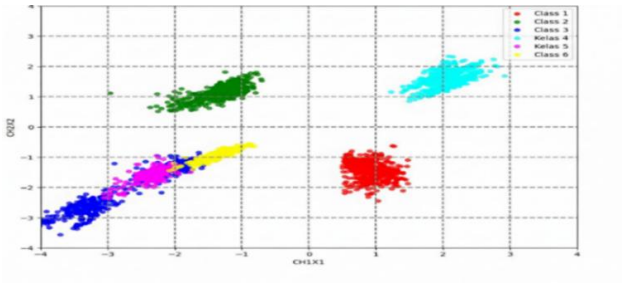


Figure 9. Data distribution of gaze and blink signals

The six-class EOG classification task was evaluated using 2,400 samples (Right: 480, Left: 480, Up: 480, Down: 480, Voluntary blink: 240, Involuntary blink: 240) and four features (Ch1 amplitude, Ch2 amplitude, Ch1 period, and Ch2 period). All models were assessed using stratified 5-fold cross-validation, and multiclass performance was reported using macro-averaged precision, recall, and F1-score to account for class imbalance.

3.2.1 Classification Performance

Table 5 summarizes the performance of KNN, SVM, and Random Forest using the same dataset, feature set, and evaluation protocol. RF achieved the highest overall accuracy (94.83%) and macro-averaged F1-score (93.66%). KNN (k=3) achieved comparable performance (accuracy 94.38%, macro-F1 93.12%) and was selected as the primary real-time classifier due to its simplicity, interpretability, and low computational overhead. Assuming 2,400 labeled samples, the estimated 95% confidence interval for the KNN accuracy is approximately 93.45% to 95.30% (binomial approximation).

Table 5. Performance Comparison of Classification Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
KNN	94.38	92.93	93.37	93.12
SVM	94.04	92.57	93.09	92.77
RF	94.83	93.48	93.89	93.66

3.2.2 ROC and Precision–Recall Analysis

To further address reviewer requests for threshold-independent evaluation, we report one-vs-rest ROC and Precision–Recall (PR) curves for the primary classifier (KNN, k=3) using stratified 5-fold cross-validation. The macro-averaged ROC-AUC was 0.983 (micro-average 0.987), and the macro-averaged PR-AUC was 0.927 (micro-average 0.958). As expected, PR performance was lower for VoluntaryBlink due to class overlap with Up gaze, which motivates additional blink-specific temporal features in future work.

Figure 10 reports the one-vs-rest ROC curves and macro-average ROC-AUC, whereas Figure 11 reports the corresponding PR curves and macro-average PR-AUC. PR analysis is included because it is

more sensitive to inter-class overlap and false-positive behavior, which is critical for distinguishing Up gaze from Voluntary Blink.

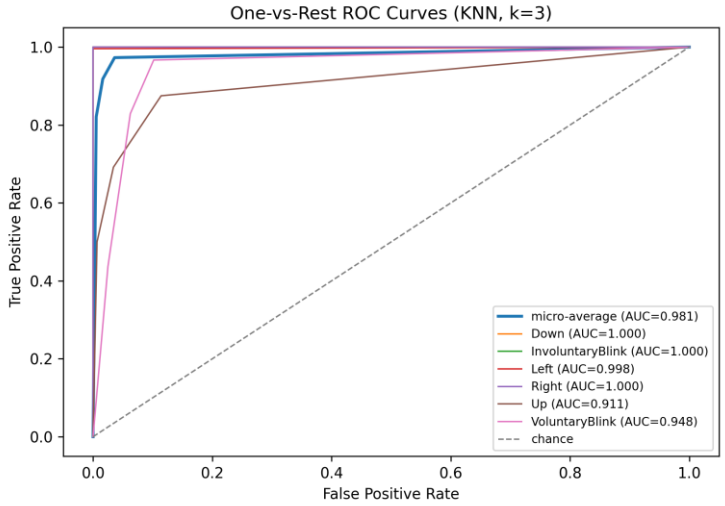


Figure 10. One-vs-rest ROC curves for EOG-based eye movement classification using KNN (k=3).

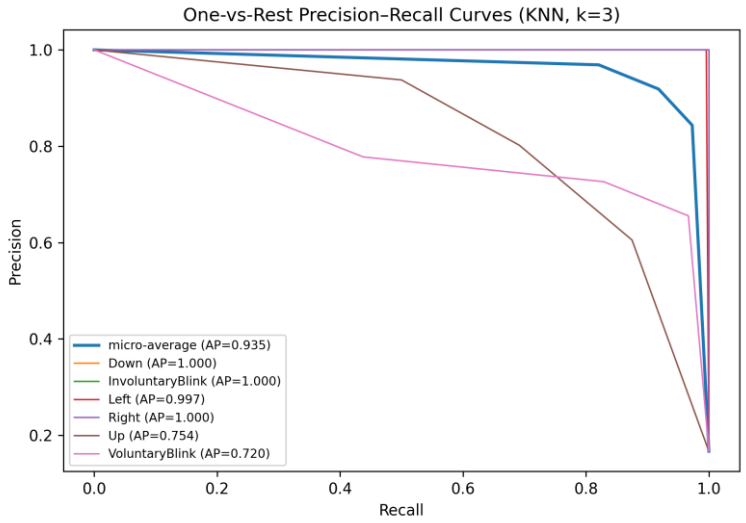


Figure 11. One-vs-rest Precision–Recall curves for EOG-based eye movement classification using KNN (k=3).

3.2.3 Ablation Study and Statistical Significance

An ablation study was conducted to quantify the contribution of each feature group in separating gaze and blink classes. Three feature configurations were evaluated under the same stratified 5-fold cross-validation protocol: (i) amplitude-only (Ch1 amplitude, Ch2 amplitude), (ii) period-only (Ch1 period, Ch2 period), and (iii) the full feature set (amplitude + period). Table 6 reports the mean performance for each configuration.

Table 6. Ablation study of feature sets for the 6-class EOG classification task (mean across stratified 5-fold CV).

Feature set	Accuracy (%)	Macro-Precision (%)	Macro-Recall (%)	Macro-F1 (%)
Amplitude-only (Ch1 amp, Ch2 amp)	91.25 ± 1.18	88.85 ± 1.50	89.20 ± 1.23	88.94 ± 1.29
Period-only (Ch1 period, Ch2 period)	50.12 ± 1.62	50.70 ± 3.83	48.54 ± 1.66	45.05 ± 1.31
Full (amp + period; 4 features)	94.38 ± 1.08	93.00 ± 1.34	93.37 ± 1.45	93.10 ± 1.33

To ensure the observed gains are not due to sampling variance, we conducted a one-sided Wilcoxon signed-rank test on fold-paired macro-F1 scores. Full vs amplitude-only yielded p=0.03125, and full vs period-only yielded p=0.03125, indicating statistically significant improvements when combining amplitude and period features.

3.2.4 Confusion Matrix Interpretation and Inter-Class Similarity

Table 7 shows the row-normalized confusion matrix (%) for the primary classifier (KNN, k=3). The model achieves very high recognition for Down and Right (both 100%), and similarly strong performance for Left and Involuntary Blink (each 99.58%), with only rare isolated confusions (e.g., Left, Involuntary Blink = 0.21% and Involuntary Blink → Voluntary Blink = 0.42%).

Misclassification is mainly concentrated between Up gaze and Voluntary Blink. Up gaze is correctly recognized at 83.96%, with the dominant confusion Up → Voluntary Blink = 15.83%, while Voluntary Blink is correctly recognized at 77.08% with the main error Voluntary Blink → Up = 22.92%. This overlap is consistent with inter-class similarity under the current features (amplitude and period), because both classes can generate short-duration, high-amplitude transients dominated by the vertical channel. Future work can improve separability by adding features such as channel peak-to-peak ratios, slope/derivative and wavelet-band energies, blink-morphology cues, and per-user adaptive thresholds to handle fatigue-related drift and minor electrode shifts.

Table 7. Confusion Matrix of EOG-Based Eye Movement Classification

Actual \ Pred	Up	Down	Right	Left	Voluntary	Involuntary
Up gaze	83.96	0	0	0	15.83	0.21
Down gaze	0	100	0	0	0	0
Right gaze	0	0	100	0	0	0
Left gaze	0.21	0	0	99.58	0	0.21
Voluntary blink	22.92	0	0	0	77.08	0
Involuntary blink	0	0	0	0	0.42	99.58

3.3 Robot Manipulator Testing

The sequential stages of the robot manipulation task in Figure 12, consisting of (a) object scanning, (b) object retrieval, (c) object positioning toward the placement target, and (d) object placement. These stages represent one complete manipulation cycle executed under EOG-based control using the optimal KNN configuration identified previously. All trials were initiated from an identical resting state to ensure consistency.

(a) (b) (c) (d)

Figure 12. (a)Scanning object, (b)Taking object, (c)Position placement object, (d)Placement object

A total of 25 manipulation trials were conducted across five object distance conditions (8, 10, 15, 17, and 20 cm), with five repeated trials per distance. Each trial comprised the full four-stage sequence shown in Fig. 12 and was considered successful when the object was released at the designated target location without grasp failure, object drop, or misplacement.

Quantitative evaluation was performed using task success rate (TSR), task completion time (CT), and positional error (PE). Across all distance conditions, the proposed system successfully completed all trials, achieving a task success rate of 100% (25/25 trials). The distance-wise quantitative performance metrics are summarized in Table 8.

Table 8. Task Performance Metrics at Different Target Distances

Trial Range	Distance (cm)	TSR (%)	Mean CT (s)	SD CT (s)	Mean PE (mm)	SD PE (mm)
1–5	8	100	12.30	0.16	3.04	0.11
6–10	10	100	13.14	0.11	3.40	0.16
11–15	15	100	14.76	0.21	4.30	0.25
16–20	17	100	16.30	0.16	5.20	0.16
21–25	20	100	18.20	0.25	6.60	0.25

As reported in Table 8, task completion time exhibited a gradual increase with increasing object distance, reflecting the longer travel distance required by the manipulator during the positioning and placement stages. Similarly, positional error showed a moderate increase at larger distances, consistent with accumulated actuation and positioning uncertainty during extended arm motion. Despite these increases, the standard deviations of both metrics remained limited across repeated trials, indicating stable and repeatable system behavior.

The results demonstrate consistent execution of the complete manipulation sequence illustrated in Figure 12, while maintaining reliable task completion and controlled placement accuracy under varying spatial conditions.

4. CONCLUSION

This study proposes an assistive teleoperation framework for a 4-DoF robotic manipulator that integrates orthogonal dual-camera vision with an electrooculography (EOG)-based human–machine interface. The results demonstrate that explicit three-dimensional target localization can be achieved using asymmetric monocular viewpoints, enabling reliable object manipulation without depth sensors or high-degree-of-freedom manipulators. Compared to previous EOG-based approaches [16], [17], the proposed system offers a simpler hardware configuration while maintaining robust spatial perception and control performance. The proposed framework demonstrates practical potential for assistive and rehabilitation applications, particularly for individuals with severe motor impairments such as quadriplegia, where intuitive and non-contact control is required for basic object retrieval tasks. Despite the promising results, this study has several limitations. The EOG command recognition model was evaluated across ten participants using a subject-dependent protocol, whereas the end-to-end robot manipulation trials were conducted with one trained operator under controlled laboratory

conditions to reduce confounding factors. In addition, the manipulation tasks were limited to basic pick-and-place operations in static environments and a restricted set of task configurations. Accordingly, the findings should be interpreted as a proof-of-concept rather than a complete validation for real-world assistive deployment. Future work will focus on expanding user evaluation for end-to-end manipulation, improving depth estimation robustness under varying lighting and occlusion, enhancing EOG class separability (especially between upward gaze and voluntary blinks), and extending system applicability to more complex and dynamic environments.

ACKNOWLEDGEMENTS

This study did not receive external funding. The authors gratefully acknowledge the Industrial Electronics Laboratory (LEI), Department of Electrical Engineering, Universitas Andalas, for providing facilities, support, and collaboration during the research.

DECLARATIONS

AI USAGE STATEMENT

Artificial intelligence–assisted tools were used for reference management and language editing, including SciSpace and Grammarly. All content was reviewed and approved by the authors, and no AI was involved in the study's conceptualization, methodology, data analysis, or conclusions. The authors take full responsibility for the originality and accuracy of the work.

AUTHOR CONTRIBUTION

The authors contributed to this research according to their respective expertise. Muhammad Ilhamdi Rusydi led the study conceptualization, methodology design, supervision, resource provision, data analysis, project administration, and manuscript drafting. Andre Paskah Gultom was responsible for software development, methodology implementation, validation, investigation, and contributed to supervision and drafting. Adam Jordan contributed to methodology refinement, algorithm and software development, and formal analysis. Rahmad Novan Nurhadi handled manuscript review and editing. Darwison provided overall supervision and ensured compliance with academic and ethical standards

CONFLICTING INTERESTS

The authors declare that there are no conflicts of interest regarding the publication of this research. The study was conducted independently, without any commercial, financial, or personal relationships that could be perceived as influencing the results or interpretations presented in this paper.

REFERENCES

[1] K. Yun, J. C. Lim, and O. Kim, "Significance of physical factors on activities of daily living in patients with tetraplegia after spinal cord injury: a retrospective study," *BMC Sports Sci. Med. Rehabil.*, vol. 16, no. 1, pp. 1–9, 2024, doi: 10.1186/s13102-024-00928-z.

[2] R. Betz *et al.*, "The 2019 revision of the International Standards for Neurological Classification of Spinal Cord Injury (ISNCSCI)—What's new?," *Spinal Cord*, vol. 57, no. 10, pp. 815–817, 2019, doi: 10.1038/s41393-019-0350-9.

[3] U. Hassan, H. Mughal, I. Mohsin, and Z. H. Khan, "Real-time control of a mobile robot using electrooculogram based eye tracking system," *5th Int. Multi-Topic ICT Conf. Technol. Futur. Gener. IMTIC 2018 - Proc.*, pp. 1–6, 2018, doi: 10.1109/IMTIC.2018.8467232.

[4] H. A. Khan *et al.*, "Design and development of machine vision robotic arm for vegetable crops in hydroponics," *Smart Agric. Technol.*, vol. 9, no. September, p. 100628, 2024, doi: 10.1016/j.atech.2024.100628.

[5] I. Rulik *et al.*, "Control of a Wheelchair-Mounted 6DOF Assistive Robot With Chin and Finger Joysticks," *Front. Robot. AI*, vol. 9, Jul. 2022, doi: 10.3389/frobt.2022.885610.

[6] Y. Mishchenko, M. Kaya, E. Ozbay, and H. Yanar, "Developing a Three- to Six-State EEG-Based Brain-Computer Interface for a Virtual Robotic Manipulator Control," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 4,

pp. 977–987, Apr. 2019, doi: 10.1109/TBME.2018.2865941.
[7] M. I. Rusydi, T. Okamoto, M. Sasaki, and S. Ito, "Line of Sight Estimation from EOG Signal with Variation of Electrode Position for Human Machine Interface," in *Proceedings of the 6th Conference of the Rehabilitation Engineering and Assistive Technology Society of Korea (RESKO)*, Jeonju, Korea, Nov. 2012, pp. 234–239.
[8] L. V. Herlant, R. M. Holladay, and S. S. Srinivasa, "Assistive teleoperation of robot arms via automatic time-optimal mode switching," *ACM/IEEE Int. Conf. Human-Robot Interact.*, vol. 2016-April, pp. 35–42, 2016, doi: 10.1109/HRI.2016.7451731.
[9] A. Francis, N. Mohan, and R. Roy, "Multi-Tasking EMG Controlled Robotic Arm," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 6, 2017, doi: 10.17148/IJARCCE.
[10] S. Crea *et al.*, "Feasibility and safety of shared EEG/EOG and vision-guided autonomous whole-arm exoskeleton control to perform activities of daily living," *Sci. Rep.*, vol. 8, no. 1, pp. 1–9, 2018, doi: 10.1038/s41598-018-29091-5.
[11] Y. Zhou *et al.*, "Shared Three-Dimensional Robotic Arm Control Based on Asynchronous BCI and Computer Vision," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 3163–3175, 2023, doi: 10.1109/TNSRE.2023.3299350.
[12] M. I. Rusydi, T. Okamoto, S. Ito, and M. Sasaki, "Controlling 3-D movement of robot manipulator using electrooculography," *Int. J. Electr. Eng. Informatics*, vol. 10, no. 1, pp. 170–185, 2018, doi: 10.15676/ijeei.2018.10.1.12.
[13] Y. Zhu, Y. Li, J. Lu, and P. Li, "A Hybrid BCI Based on SSVEP and EOG for Robotic Arm Control," *Front. Neurorobot.*, vol. 14, Nov. 2020, doi: 10.3389/fnbot.2020.583641.
[14] M. S. Amri bin Suhaimi, K. Matsushita, T. Kitamura, P. W. Laksono, and M. Sasaki, "Object Grasp Control of a 3D Robot Arm by Combining EOG Gaze Estimation and Camera-Based Object Recognition," *Biomimetics*, vol. 8, no. 2, Jun. 2023, doi: 10.3390/biomimetics8020208.
[15] A. Anandika, P. D. Laksono, M. S. A. bin Suhaimi, J. Muguro, and M. I. Rusydi, "Enhancing Interface Efficiency: Adaptive Virtual Keyboard Minimizing Keystrokes in Electrooculography-Based Control," *J. Nas. Tek. Elektro*, pp. 64–72, Dec. 2023, doi: 10.25077/jnte.v12n3.1160.2023.
[16] M. I. Rusydi, Y. Mori, T. Okamoto, M. Sasaki, and S. Ito, "Development of an Eog Based Robot Manipulator and End Point Direction Control System," *J. Japan Soc. Appl. Electromagn. Mech.*, vol. 22, no. 2, pp. 293–299, 2014, doi: 10.14243/jsaem.22.293.
[17] Y. Mishchenko, M. Kaya, E. Ozbay, and H. Yanar, "Developing a Three- to Six-State EEG-Based Brain-Computer Interface for a Virtual Robotic Manipulator Control," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 4, pp. 977–987, 2019, doi: 10.1109/TBME.2018.2865941.
[18] M. I. Rusydi, T. Okamoto, S. Ito, and M. Sasaki, "Controlling 3-D Movement of Robot Manipulator using Electrooculography," vol. 10, no. 1, pp. 170–186, 2018, doi: 10.15676/ijeei.2018.10.1.12.
[19] M. I. Rusydi, M. Sasaki, and S. Ito, "Affine transform to reform pixel coordinates of EOG signals for controlling robot manipulators using gaze motions," *Sensors (Switzerland)*, vol. 14, no. 6, pp. 10107–10123, 2014, doi: 10.3390/s140610107.
[20] M. Anschober, R. Edlinger, R. Froschauer, and A. Nüchter, "Inverse Kinematics of an Anthropomorphic 6R Robot Manipulator Based on a Simple Geometric Approach for Embedded Systems," *Robotics*, vol. 12, no. 4, Aug. 2023, doi: 10.3390/robotics12040101.
[21] R. K. Megalingam, V. Shriram, B. Likhith, G. Rajesh, and S. Ghanta, "Monocular distance estimation using pinhole camera approximation to avoid vehicle crash and back-over accidents," *Proc. 10th Int. Conf. Intell. Syst. Control. ISCO 2016*, 2016, doi: 10.1109/ISCO.2016.7727017.
[22] Z. Zhang, R. Zhao, E. Liu, K. Yan, and Y. Ma, "A single-image linear calibration method for," *Measurement*, no. July, 2018, doi: 10.1016/j.measurement.2018.07.085.
[23] Z. Zhang, "A Flexible New Technique for Camera Calibration," vol. 1998, 2008.
[24] M. I. Rusydi *et al.*, "Electrooculography signal as alternative method to operate wheelchair based on SVM classifier," in *3rd Conference on Innovation in Technology and Engineering Science 2022 (CITES2022) — AIP Conference Proceedings*, AIP Publishing, 2024. doi: 10.1063/5.0200941.
[25] M. I. Rusydi, T. Okamoto, S. Ito, and M. Sasaki, "Controlling 3-D movement of robot manipulator using electrooculography," *Int. J. Electr. Eng. Informatics*, vol. 10, no. 1, pp. 170–185, Mar. 2018, doi: 10.15676/ijeei.2018.10.1.12.
[26] A. Papetti, M. Ciccarelli, A. Manni, A. Caroppo, and G. Rescio, "Investigating the Use of Electrooculography Sensors to Detect Stress During Working Activities," *Sensors*, vol. 25, no. 10, May 2025, doi: 10.3390/s25103015.