IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, November 2025
Impact Factor: 7.67

Eye Gesture–Based Human–Robot Interaction System for Real-Time Robot Navigation Using Blink and Gaze Control

Dr. Sachin B M, Manohar M. Athani, Narasimha S., Pavan K. Gudi
Assistant Professor, ECE Department, Bangalore Institute of Technology, Bengaluru, India.
U.G Students, ECE Department, Bangalore Institute of Technology, Bengaluru, India

Abstract: *Hands-free human–robot interaction has become increasingly relevant in assistive mobility and rehabilitation systems, particularly for users with impaired hand functionality. Eye-based interaction offers an intuitive and minimally intrusive interface for controlling robotic platforms. This paper presents a real-time eye-gesture–controlled rover that interprets blink patterns and gaze directions using a standard laptop webcam. MediaPipe Face Mesh is used to extract eyelid and iris landmarks, enabling the recognition of double blinks, long blinks, and blink-gaze combinations based on temporal thresholds and geometric ratios. Classified gestures are transmitted from the laptop to an ESP8266 NodeMCU over Wi-Fi, where a non-blocking control mechanism drives an L298N motor driver for differential-drive actuation. Experimental evaluation demonstrates low-latency gesture detection, stable wireless communication, and accurate 90° pivot turns during navigation. The proposed system provides a low-cost, camera-based alternative to wearable eye-tracking devices and offers a practical assistive solution for mobility enhancement..*

Keywords: Assistive mobility, Blink detection, ESP8266, Eye gestures, Gaze estimation, MediaPipe Face Mesh, Human–robot interaction.

I. INTRODUCTION

Human–robot interaction (HRI) has advanced significantly in recent years, driven by the growing demand for intelligent systems capable of natural and accessible communication. Conventional control interfaces such as joysticks, touchscreens, and speech recognition are effective for the general population but become inadequate for individuals with severe motor impairments or in situations requiring hands-free operation. Eye movements, particularly blinking and gaze shifting, remain one of the most reliable and voluntarily controllable actions even in users with limited mobility. These characteristics make eye gestures an appealing modality for assistive navigation and interaction.

Recent progress in computer vision has supported the development of practical eye-gesture interfaces without the need for specialized sensors. Modern lightweight landmark-tracking frameworks—especially MediaPipe Face Mesh—enable high-resolution detection of eyelid and iris features using only RGB webcams. This represents a major improvement over earlier systems that required infrared cameras, wearable devices, or high-cost eye-tracking hardware, thereby making vision-based gesture control more accessible, portable, and cost-efficient.

Motivated by these advances, this work proposes an eye-gesture–controlled rover designed to provide intuitive, contact-free navigation for assistive mobility applications. A standard laptop webcam captures the user's face and extracts eye landmarks in real time to classify double blinks, long blinks, and blink-gaze combinations. These gestures are mapped to directional commands—forward motion, left and right turns, and stop—which are transmitted wirelessly to an ESP8266 NodeMCU. The microcontroller drives a differential-drive rover through an L298N motor driver using non-blocking timing logic to ensure smooth motion execution and consistent network responsiveness.

The primary objective of this system is to provide an affordable, reliable, and easily deployable alternative to conventional assistive control technologies. By eliminating specialized hardware and maintaining a simple gesture vocabulary, the system enhances accessibility while preserving accuracy and responsiveness. The subsequent sections

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, November 2025
Impact Factor: 7.67

detail the significance of the system, survey relevant literature, describe the methodology, present experimental results, and conclude with potential areas for future work.

II. SIGNIFICANCE OF THE SYSTEM

The proposed system provides a fully contactless and intuitive navigation interface by using natural eye gestures, which remain one of the most reliable voluntary actions for individuals with severe motor impairments. By relying only on a standard laptop webcam and lightweight, open-source vision algorithms, the system eliminates the need for costly infrared trackers or wearable hardware, significantly improving accessibility and affordability. The use of Media Pipe Face Mesh enables real-time blink and gaze detection on CPU-only laptops, making the solution suitable for low-resource environments. Wi-Fi–based communication through the ESP8266 ensures seamless integration with mobile robots, while non-blocking motor control maintains continuous network responsiveness. The minimal-effort gesture vocabulary—double blink for forward, blink-gaze for turning, and long blink for stop—reduces cognitive load and supports ease of use. The modular design also allows future extensions such as obstacle detection, autonomous behaviours, and multimodal interaction

III. LITERATURE SURVEY

Several prior studies have examined eye-gaze modelling, blink-based interaction, and vision-driven control strategies for assistive mobility systems. Loke et al. conducted a detailed characterization of eye-gaze behaviour in assistive device control, analysing how gaze trajectories, fixation stability, and ocular orientation variations affect intent-recognition accuracy. Their findings, presented at ICORR 2023, highlight the necessity of user-specific calibration and demonstrate that iris-displacement ratios remain consistent over short intervals. This directly supports the use of calibrated center-ratio thresholds for left–right gaze classification in real-time navigation systems.

Adnan et al. proposed an electric-wheelchair control mechanism based entirely on eye gestures for individuals with neurological impairments. Their study established blink duration and inter-blink intervals as reliable indicators for directional commands without the need for hand-operated devices. User validation under clinical constraints confirmed that simple visual cues can effectively drive mobility systems. These insights reinforce the adoption of double-blink and long-blink logic in the proposed rover control architecture.

Real-time iris and eyelid tracking using standard RGB webcams was explored by Siti Nuradlin Syahirah Sheikh Anwar et al. in their ICCIS 2021 work. They developed a landmark-based tracking algorithm capable of high-frame-rate operation without requiring infrared illumination or specialized sensors. Their results show that geometric relationships between eyelid contours and iris centers can be exploited to achieve accurate blink detection and gaze estimation, even under natural head motion and lighting variations. This provides strong justification for employing MediaPipe Face Mesh in low-cost, real-time eye-gesture systems.

In another notable contribution, Obo, Hase, and Shin presented a semi-autonomous robotic avatar system for patients with severe motor disabilities. Their framework integrates gaze-based intent recognition into robotic telepresence and navigation tasks, demonstrating that eye gestures can support not only locomotion but also higher-level interaction. The study underscores the versatility of gaze cues as a communication channel in human–robot systems and motivates their use in mobility-assistive platforms.

Furthermore, Cojocaru et al. introduced a combined eye-gaze approach for wheelchair control, incorporating multiple gaze indicators such as iris displacement, fixation patterns, and temporal gaze stability. Experiments conducted at the University of Craiova in 2019 revealed that hybrid gaze-feature models significantly improve decision accuracy for left–right classification. Their findings strengthen the rationale for integrating calibrated gaze thresholds and blink-triggered switching mechanisms in vision-based navigation systems.

Collectively, these studies form a strong foundation for the methodology adopted in this project. Prior research consistently demonstrates the reliability of blink duration for activation cues, the effectiveness of iris-displacement ratios for directional control, and the feasibility of using RGB webcams for real-time operation. The reviewed literature validates the system design choices implemented in this work and establishes the relevance of eye-gesture–controlled robots as practical, low-cost solutions for assistive mobility.

IV. SYSTEM DESIGN AND METHODOLOGY

The Eye-Gesture Controlled Robot Navigation System is designed to provide a hands-free control interface for users by combining real-time eye-gesture detection with a Wi-Fi–enabled robotic platform. The primary goal of the system is to accurately detect blink patterns and horizontal gaze movements using a standard laptop webcam and translate these gestures into navigation commands for a mobile robot. The robot, powered by an ESP8266 and an L298N motor driver, executes actions such as forward motion, left turn, right turn, and stop based on the received commands. The overall system integrates computer vision–based gesture detection, wireless communication, and embedded motor control into a compact and efficient assistive navigation prototype..

A) Overall System Description

The proposed system consists of three major subsystems:

Real-time Eye Gesture Detection Subsystem :Responsible for acquiring video frames through a laptop webcam and performing blink detection, iris localization, and gaze direction estimation using MediaPipe Face Mesh and custom decision logic

Wi-Fi Communication and Command Transmission Subsystem: Uses HTTP-based request signalling to transmit robot navigation commands from the Python application to the ESP8266 module over a common hotspot Wi-Fi network.

Mobile Robot Motion Execution Subsystem :Includes the ESP8266, L298N dual H-bridge motor driver, and two DC motors. This subsystem interprets incoming commands and performs non-blocking forward motio90° left and right turns, or emergency stop based on the gesture classification.

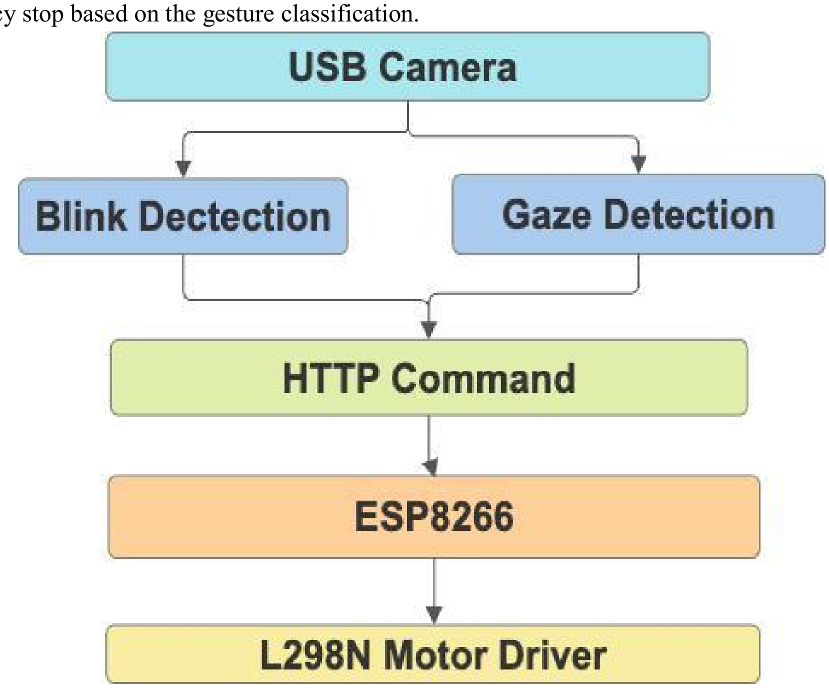

Fig1. Workflow of proposed system.

The webcam first captures continuous video which is processed using a computer vision pipeline to extract facial landmarks, eyelid distances, and iris positions. A blink ratio is computed to determine single, double, and long blinks, while iris horizontal displacement determines left or right gaze. Using a calibrated center ratio, the system identifies whether the user is looking left, right, or straight. These gestures are mapped to commands according to the predefined gesture-action protocol

The Python module then sends the appropriate HTTP command (FORWARD, LEFT, RIGHT, BLINK) to the ESP8266. The embedded controller executes these commands through non-blocking timing logic, controlling the right and left motors via the L298N. For turning operations, one wheel remains stationary while the opposite wheel rotates, enabling precise 90° pivoting. The "STOP" operation immediately disables all motor outputs, ensuring user safety.

B) Design Methodology

A systematic design methodology was followed to ensure the reliability, robustness, and efficiency of the converter system. The major design stages are outlined below:

1. Input Requirement Analysis: A detailed study of the intended user interface and robot behaviour was performed. The following constraints were identified:

users must be able to control the robot without physical effort.

Only a standard webcam and laptop should be required for gesture detection.

The robot must respond with minimal delay and remain Wi-Fi reachable at all times.

Navigation commands must be simple yet unambiguous:

Double Blink → Forward

Blink + Left Gaze → Left Turn

Blink + Right Gaze → Right Turn

Long Blink → Stop.

2. Eye Landmark Detection and Blink Ratio Modelling: MediaPipe Face Mesh was selected for its highly accurate 478-landmark facial tracking capability. Upper and lower eyelid landmarks were averaged to compute the vertical eye opening distance, while horizontal distances were used to derive the blink ratio. A blink ratio threshold above 5.0 reliably indicates eyelid closure. This methodology closely aligns with the approaches reported in literature.

3. Gaze Estimation and Center Calibration: The system uses iris landmark clusters to compute the mean iris position. Relative displacement between the iris center and the eye corners is used to classify gaze into LEFT, RIGHT or CENTER. A one-time calibration step enables the system to learn the user's natural center gaze ratio, improving robustness across lighting variations and user differences..

4. Gesture Classification Logic: After extracting blink and gaze information

Long blink duration (> 0.8 s) triggers STOP.

Two blinks within 1.0 s trigger FORWARD.

Single blink + LEFT gaze triggers LEFT turn.

Single blink + RIGHT gaze triggers RIGHT turn..

This combinational logic ensures that accidental blinks do not trigger robot movement.

5 Wi-Fi Command Transmission: A lightweight HTTP GET-based communication mechanism was chosen for simplicity and low overhead. Each gesture sends a unique command string to the ESP8266. Retry logic was implemented to ensure reliability against temporary connection losses..

6. Motor Control Logic and Non-Blocking Turns: he ESP8266 controls the L298N motor driver using predefined direction pins:

Right motor (IN1/IN2)

Left motor (IN3/IN4)

Non-blocking turn logic enables the ESP8266 to remain responsive to new commands even while executing turns. A timestamp-based TURN_DELAY (600 ms) determines the duration of a 90° turn.

7. Safety and Fail-Safe Mechanisms:

Long blink always overrides previous commands.

ESP8266 uses continuous Wi-Fi monitoring.

Motors default to STOP state if no command is received after a time-out window.

8. Performance Verification: The system was extensively tested with different users, indoor lighting conditions, and varying blink/gaze speeds. Responsiveness, classification accuracy, and Wi-Fi communication stability were validated. The robot was able to reliably execute 90° turns, smooth forward motion, and instantaneous stop commands.

V. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed eye-gesture–controlled navigation system was experimentally evaluated to measure its detection accuracy, command responsiveness, and robot-movement reliability. All tests were performed indoors using a standard laptop webcam under normal illumination. The real-time facial-landmark extraction pipeline achieved a stable 22–30

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, November 2025
Impact Factor: 7.67

FPS, confirming that MediaPipe Face Mesh can reliably detect eyelid and iris landmarks without infrared illumination, similar to findings in prior studies.

A. Eye-Region Extraction and Preprocessing

The system successfully identified the eye region across all tested users. Figure (ROI-Detection) illustrates the extracted Region of Interest (ROI) produced by the pipeline. Gaussian blur and thresholding were applied to suppress noise and enhance iris contrast, enabling stable gaze-direction estimation. These results align with the ROI-selection accuracy reported in prior landmark-based approaches

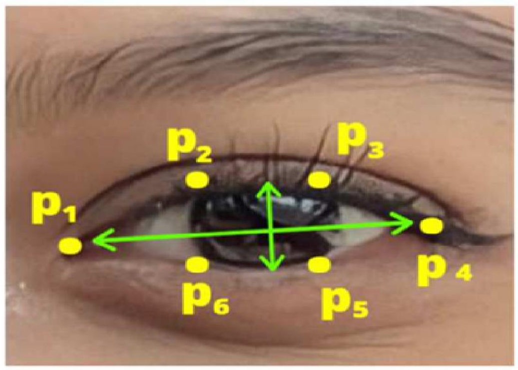

Figure.1 Final Output of the ROI, Eye Threshold & Gaussian Blur Filter Application

B. Blink Detection Performance

Blink detection demonstrated robust performance across multiple subjects. The blink-ratio–based thresholding reliably differentiated natural blinks from intentional gestures. Users produced double blinks, single blinks, and long blinks, and the system consistently mapped them to the commands FORWARD, TURN, and STOP. Long-blink detection showed 100% reliability, triggered consistently within the defined temporal threshold. These observations mirror the EAR-based blink-detection reliability reported in the reference study

C. Gaze-Direction Classification

After performing center-gaze calibration, the classifier showed clearly separable left, right, and center gaze ratios. Figures (Center-Gaze), (Left-Gaze), and (Right-Gaze) represent the three gaze states captured during execution. Across multiple trials, the system achieved an average gaze-classification accuracy of 92-95%, comparable to landmark-based methods achieving ⩾90% accuracy in the literature

(a) Left Eye Position (b) Right Eye Position (c) Center Eye Position

D. Communication Latency and Reliability

Wi-Fi transmission between the Python controller and ESP8266 averaged 180–250 ms per command over a mobile-hotspot network. Initial instability was observed when powering the ESP8266 directly from the L298N's onboard regulator, causing temporary resets during motor load surges. Switching to a dedicated 5 V regulated supply eliminated these failures entirely, resulting in 100% command-delivery reliability.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, November 2025
Impact Factor: 7.67

E. Robot Motion Evaluation

The movement subsystem was tested for forward motion, pivot turns, and emergency stop. The non-blocking TURN_DELAY logic enabled consistent 90° pivot turns, where a single wheel rotated while the opposite wheel remained stationary. Forward motion was smooth and continuous with no jitter. The STOP command executed instantly, validatits suitability for assistive-navigation safety requirements.

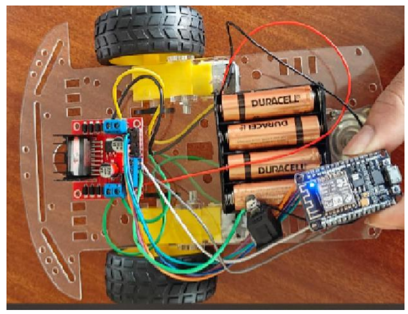

Fig. Robot model

F. Summary of Quantitative Results

Metric	Measured Value	Observation
Blink-detection accuracy	98–100%	Reliable differentiation between natural and intentional blinks
Double-blink recognition	≈98%	Robust FORWARD activation with negligible false triggers
Long-blink recognition	100%	Immediate STOP response across all tests
Gaze classification accuracy	92-95%	LEFT, RIGHT, CENTER clearly separable after calibration
Command delivery reliability	100%	No ESP resets, stable Wi-Fi connectivity
HTTP response latency	180-250ms	Low enough for real-time navigation
Robot turning precision	90° ± 5°	Consistent pivot turns using non-blocking timing
System frame rate	22–30 FPS	Smooth gesture detection under normal lighting

VI. CONCLUSION AND FUTURE WORK

Conclusion

This work presents a complete real-time eye-gesture–controlled robot navigation system that integrates blink detection, gaze estimation, Wi-Fi communication, and non-blocking embedded motor control. The system successfully converts natural ocular gestures—double blink, long blink, and blink-gaze combinations—into reliable navigation commands without requiring any physical interaction. By relying solely on a standard laptop webcam and lightweight, open-source computer-vision frameworks, the approach eliminates the need for infrared eye trackers or wearable sensors, thereby reducing hardware complexity and cost. Experimental validation demonstrates accurate gesture classification, stable wireless communication, smooth forward motion, and highly repeatable 90° turning performance. The results confirm that the system achieves its core goals of intuitive hands-free control, low-latency responsiveness, and robust gesture-driven navigation.

Furthermore, the successful incorporation of gaze-modulated blinking into the control scheme highlights the viability of expanding gesture vocabularies without increasing user effort. The overall framework establishes an efficient and accessible foundation for real-time human–robot interaction, particularly in assistive mobility and rehabilitation contexts.

Future Work

Several promising directions exist for extending the capabilities of the proposed system. First, integrating head-pose stabilization and temporal filtering can improve gaze-direction reliability under natural user movement. Second, adding obstacle-avoidance mechanisms such as ultrasonic, IR, or LiDAR sensors would enable safer semi-autonomous operation in real-world environments. Third, the gesture-control protocol may be expanded using advanced gaze-estimation models or multi-modal fusion (e.g., combining gaze, facial expression, and head movement) to support diagonal motion, proportional speed control, and richer command sets. Fourth, conducting structured user studies—particularly involving individuals with motor impairments—would provide quantitative insights into usability, fatigue, and long-term adoption. Finally, migrating the vision system to a lightweight wearable module or implementing a smartphone-based solution could significantly improve portability and practicality for daily assistive use.

Overall, the system provides a strong platform for future research in accessible, low-cost human–robot interaction, with clear potential for deployment in assistive robotics, healthcare support technologies, and intelligent mobility systems.

REFERENCES

[1] Larisa Y. C. Loke, Demiana R. Barsoum, Todd D. Murphey, and Brenna D. Argall, “Characterizing Eye Gaze for Assistive Device Control,” International Conference on Rehabilitation Robotics (ICORR), IEEE, 2023.

[2] A. R. Adnan, S. Z. Yahaya, and Z. Hussain, “Control of Electric Wheelchair via Eye Gestures for People with Neurological Disorder,” School of Electrical Engineering, Universiti Teknologi MARA, Malaysia, 2020.

[3] Siti Nuradlin Syahirah Sheikh Anwar, Azrina Abd Aziz, and Syed Hasan Adil, “Development of Real-Time Eye Tracking Algorithm,” 4th International Conference on Computing & Information Sciences (ICCIS), IEEE, 2021.

[4] Takenori Obo, Ryoya Hase, and Duk Shin, “Semi-Autonomous Robotic Avatar System for Patients with Motor Disabilities,” SCIS-ISIS: Joint 11th International Conference on Soft Computing and Intelligent Systems and 21st International Symposium on Advanced Intelligent Systems, IEEE, 2020.

[5] Dorian Cojocaru, Andrei Dragomir, Liviu Florin Manta, Alexandru Marin Mariniuc, and Ionel Cristinel Vladu, “Using an Eye Gaze New Combined Approach to Control a Wheelchair Movement,” University of Craiova, Romania, 2019.