IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 1, January 2026

DOI: 10.17148/IJARCCE.2026.15182

Air Command : A Vision Driven Gesture and Gaze Control System

Avinash Gowda S[1], Usha M[2]

Department of MCA, BIT, K.R. Road, V.V. Pura, Bangalore, India

Abstract: Human computer interaction is commonly performed using physical input devices, which may not be suitable in all environments or for all users. This paper presents Air Command, a vision-driven gesture and gaze control system that enables hands-free computer interaction using a standard webcam. The system combines gaze-based cursor movement with hand gesture-based command execution, allowing users to perform operations such as clicking, scrolling, zooming, and application control without physical contact. Eye blink detection is used for mouse click actions, while predefined hand gestures trigger system commands. The proposed approach operates in real time on consumer-grade hardware and incorporates stability mechanisms to reduce unintended actions. Experimental results demonstrate reliable performance under normal conditions, highlighting the feasibility of vision-based interaction as an effective alternative to traditional input devices.

Keywords: Human–Computer Interaction, Computer Vision, Gaze Tracking, Hand Gesture Recognition, Touch-Free Interaction, Assistive Computing.

I. INTRODUCTION

Human–computer interaction traditionally relies on physical input devices such as keyboards and mice. While these devices are effective in standard computing environments, they may not be suitable in situations that require touch-free operation or improved accessibility. Users with physical limitations and environments that demand hygienic interaction often face challenges when using conventional input methods.

With recent progress in computer vision, vision-based interaction techniques have emerged as an alternative to physical input devices. These techniques enable users to interact with computers using natural actions such as eye movement and hand gestures captured through a camera. However, many existing systems struggle with real-time stability, accuracy, and usability when deployed using low-cost hardware.

This paper presents Air Command, a vision-driven gesture and gaze control system designed to provide reliable touch-free computer interaction using a standard webcam. The system combines gaze-based cursor movement and blink-based clicking with hand gesture-based command execution. By clearly separating cursor control and command actions and incorporating stability mechanisms, the proposed system offers a practical and cost-effective solution for touch-free human–computer interaction.

1.1 Project Description

The Air Command: A Vision-Driven Gesture and Gaze Control System is designed to enable touch-free computer interaction using visual cues captured through a standard webcam. The system allows users to control cursor movement using eye gaze, perform mouse click actions through deliberate eye blinks, and execute system commands using predefined hand gestures. By relying solely on camera-based input, the system eliminates the need for physical input devices or specialised hardware, making it suitable for accessibility-focused and touch-free computing environments.

The project adopts a modular architecture that separates continuous cursor control from discrete command execution to improve interaction clarity and stability. Gaze-based input is dedicated to pointer navigation and clicking, while hand gestures are mapped to actions such as scrolling, zooming, application launching, and media control. Stability-enhancing mechanisms are incorporated to reduce unintended actions and improve real-time performance. Overall, the system provides a practical, low-cost, and user-friendly alternative to traditional human–computer interaction methods.

1.2 Motivation

The motivation for this project arises from the limitations of traditional computer input devices in scenarios where physical interaction is inconvenient or inaccessible. Users with motor impairments, repetitive strain issues, or those

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 15, Issue 1, January 2026

DOI: 10.17148/IJARCCE.2026.15182

operating in hygienic or touch-restricted environments often face challenges when relying on keyboards and mice. Existing touch-free interaction solutions frequently depend on expensive hardware or lack the stability required for continuous everyday use, limiting their practical adoption.

Advancements in computer vision provide an opportunity to develop cost-effective, camera-based interaction systems that use natural human actions such as eye movement and hand gestures. The motivation behind the Air Command system is to leverage these advancements to create an intuitive and accessible interaction model that operates using a standard webcam. By combining gaze-driven cursor control with gesture-based command execution, the project aims to deliver a reliable, touch-free alternative that enhances usability, accessibility, and user comfort in real-world computing environments.

II. RELATED WORK

Paper [1] investigates multimodal human–computer interaction by combining eye gaze and hand gestures for touch-free control. The study demonstrates that using gaze for pointer navigation and gestures for command execution improves interaction speed and accuracy. However, the system relies on controlled environments and does not fully address stability issues caused by involuntary eye movements.

Paper [2] explores camera-based eye gaze tracking techniques for cursor control using appearance-based models. While the approach reduces dependency on specialised hardware, the study highlights challenges such as cursor jitter, sensitivity to lighting variations, and the need for frequent recalibration, which affect long-term usability.

Paper [3] focuses on vision-based hand gesture recognition for system control and navigation. The results show effective gesture detection for predefined actions; however, the system lacks robust validation mechanisms such as gesture hold-time or cooldown logic, leading to accidental command execution during natural hand motion.

Paper [4] examines hybrid interaction systems that integrate gaze tracking and hand gestures to enhance flexibility. Although the multimodal approach improves interaction capability, the study reports issues related to unclear separation between cursor control and command execution, increasing cognitive load on users.

Paper [5] reviews vision-based accessibility interfaces developed for users with motor impairments. The survey emphasizes the importance of low-cost, camera-driven solutions and real-time feedback. It concludes that practical adoption requires systems that are stable, easy to deploy, and suitable for everyday computing environments.

III. METHODOLOGY

A. System Environment

The system environment is designed to evaluate the Air Command: A Vision-Driven Gesture and Gaze Control System under realistic and practical usage conditions. The application operates as a standalone desktop-based system where a single user interacts with the computer through a standard webcam. The webcam continuously captures real-time video frames containing facial and hand movements, which serve as the primary input for interaction.

The processing environment consists of a Python-based backend responsible for video frame acquisition, feature extraction, and interaction control. Computer vision modules analyse facial landmarks, eye movement, and hand positions to interpret gaze direction, blink patterns, and hand gestures. These inputs are processed locally on the user's machine without transmitting data to external servers, ensuring user privacy and low-latency response. System-level commands such as cursor movement, clicking, scrolling, and application control are executed directly through the operating system interface.

This environment simulates a real-world touch-free computing scenario where users can perform everyday tasks without relying on physical input devices. By operating entirely on consumer-grade hardware and a standard webcam, the system ensures ease of deployment, reliable real-time performance, and suitability for accessibility-focused and touch-free interaction environments.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ∺ Peer-reviewed & Refereed journal ∺ Vol. 15, Issue 1, January 2026

DOI: 10.17148/IJARCCE.2026.15182

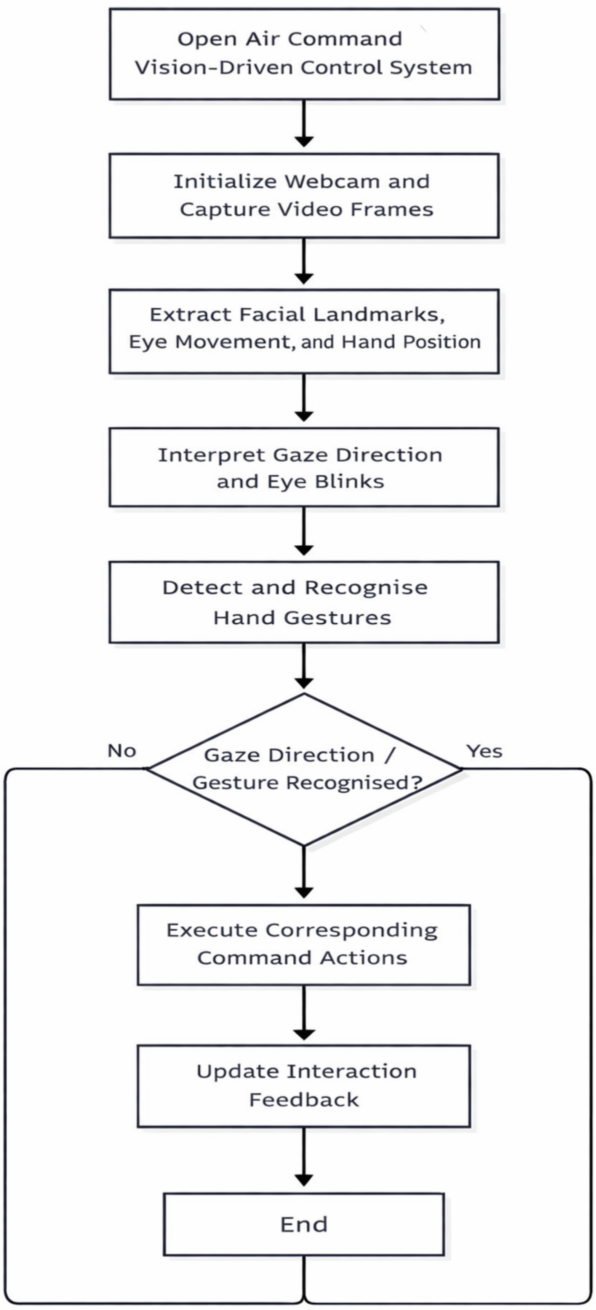

Fig.1.Flowchart of methodology

B. Vision-Based Interaction Architecture

- Client-Side Processing: In the Air Command system, real-time video input is captured using a standard webcam connected to the user's computer. The video stream is processed locally to extract visual features related to facial orientation, eye movement, blink patterns, and hand positions. Frame-level preprocessing is performed to reduce noise, handle lighting variations, and ensure stable detection before interaction logic is applied.

- Interaction Processing: The processed visual features are analysed using computer vision algorithms to determine gaze direction and recognise predefined hand gestures. Gaze information is used to control cursor movement, while intentional eye blinks trigger mouse click actions. Hand gesture recognition is used for executing system commands such as scrolling, zooming, application launching, and media control. All interaction processing is performed locally to ensure low latency and user privacy.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 1, January 2026
DOI: 10.17148/IJARCCE.2026.15182

C. Adaptive Interaction Mechanism

The interaction logic of the system is designed to be adaptive and extensible. Parameters such as cursor sensitivity, dead-zone thresholds, gesture hold-time, and cooldown intervals can be adjusted to suit different users and usage conditions. This adaptability allows the system to maintain stable performance across variations in user behaviour, camera positioning, and environmental conditions. The modular design also supports future enhancement by enabling the addition of new gestures and interaction rules without affecting existing functionality.

D. Implementation Flow

1. The user starts the Air Command system on the computer.
2. The webcam is initialised and begins capturing real-time video frames.
3. Facial landmarks and eye regions are detected from the video stream.
4. Gaze direction is estimated and mapped to cursor movement.
5. Eye blink patterns are analysed to perform mouse click actions.
6. Hand regions are detected and finger positions are analysed.
7. Recognised gestures are validated using hold-time and cooldown logic.
8. Valid gestures are mapped to corresponding system commands.
9. The system continues interaction until the user issues an exit command.

E. Hardware and Software Requirements

- **Hardware:** A standard computer system with a minimum of 8 GB RAM and a built-in or external webcam is sufficient to run the Air Command system. No specialised sensors or external hardware devices are required, making the system suitable for consumer-grade environments.

- **Software:** Python 3.8+ is used for system development and interaction logic. Computer vision processing is implemented using OpenCV and MediaPipe libraries. PyAutoGUI is used for executing system-level actions such as cursor movement and keyboard input. The system is designed to run on common operating systems without requiring additional software dependencies.

IV. SIMULATION AND EVALUATION FRAMEWORK

This section describes the system design, execution flow, and evaluation strategy adopted for the Air Command: A Vision-Driven Gesture and Gaze Control System. The framework focuses on validating the effectiveness of vision-based interaction techniques under realistic usage conditions. The system is implemented using Python as the core development platform, integrating real-time computer vision modules to enable gaze-driven cursor control and hand gesture-based command execution using a standard webcam.

A. System Architecture and Workflow

The overall architecture is designed to provide stable, responsive, and touch-free computer interaction while maintaining usability and real-time performance. The key components of the system are outlined below:

- User Interaction Layer: The user interacts with the system through natural visual actions captured by a webcam. Eye movement, blink patterns, and hand gestures act as the primary input mechanisms, eliminating the need for physical input devices.
- Vision Processing Layer: This layer processes real-time video frames to detect facial landmarks, eye regions, and hand positions. Gaze direction, blink events, and gesture features are extracted and analysed to interpret user intent accurately.
- Interaction Control Module: The processed visual inputs are mapped to corresponding system actions. Gaze input controls cursor movement, blink detection triggers mouse click actions, and validated hand gestures execute commands such as scrolling, zooming, application control, and navigation. Stability mechanisms such as dead-zone handling, hold-time validation, and cooldown logic are applied to prevent unintended actions.

B. Simulation Setup

The simulation environment is designed to replicate real-world usage of the Air Command system across different users and operating conditions.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 1, January 2026
DOI: 10.17148/IJARCCE.2026.15182

- **User Interaction Simulation:** Multiple test scenarios involving varied gaze directions, blink patterns, and hand gestures are evaluated to assess system responsiveness, accuracy, and interaction stability during continuous use.
- **Scenario Testing:** Different scenarios such as neutral gaze positioning, directional cursor movement, valid and invalid gesture execution, prolonged usage, and exit command handling are tested to ensure robustness and reliable system behaviour.

C. Interaction and Evaluation Process

During simulation, real-time video input is continuously processed through the vision pipeline. Detected gaze direction and gestures are validated and translated into corresponding system actions. The interaction results are observed across repeated test cases to evaluate consistency, responsiveness, and accuracy of cursor movement and command execution under varying conditions.

D. Results and Observations

- **Interaction Accuracy:** The system demonstrated accurate interpretation of gaze direction and hand gestures, enabling smooth cursor control and reliable execution of predefined commands.
- **System Stability:** The integration of vision processing and interaction control modules functioned effectively, with minimal latency and reduced unintended actions during continuous operation.
- **Usability and Practicality:** The evaluation confirmed that the system is intuitive for users with minimal training and suitable for everyday touch-free computer interaction, highlighting its potential for accessibility-focused and real-world deployment.

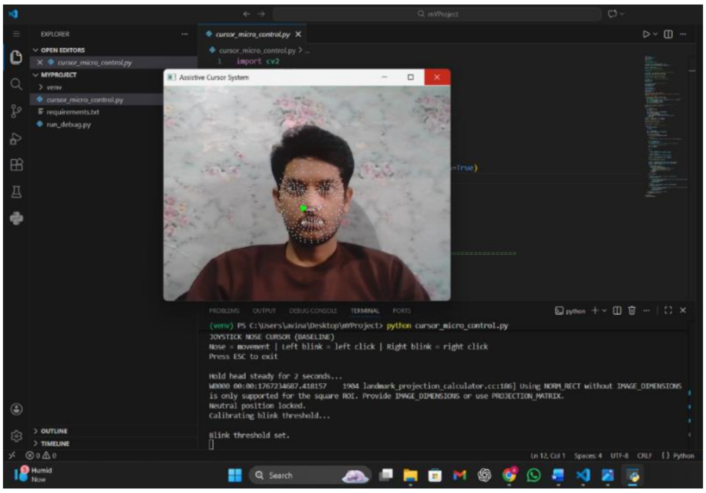

Fig .2. Eye Gaze Mouse Control

Interaction Performance and Adaptability Analysis

- **System Stability and Responsiveness:** The Air Command system demonstrated stable real-time performance during continuous operation. Gaze-based cursor movement and hand gesture recognition remained consistent across repeated interaction cycles without noticeable degradation. The integration of dead-zone handling and smoothing logic ensured reliable cursor behaviour even during prolonged usage.

- **Interaction Accuracy Improvement:** The accuracy of cursor control and gesture recognition improved with calibrated sensitivity parameters and gesture hold-time validation. The system effectively distinguished intentional eye blinks and deliberate hand gestures from natural movements, confirming the suitability of the selected vision-based features for touch-free interaction.

- **Handling of User and Environmental Variations:** The system successfully handled variations in user behaviour, including differences in head movement, eye positioning, hand size, and interaction speed. It also

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 15, Issue 1, January 2026

DOI: 10.17148/IJARCCE.2026.15182

adapted well to moderate changes in lighting conditions and camera placement, maintaining accurate gaze estimation and gesture detection across diverse usage scenarios.

- **Action Interpretability and Feedback:** Detected gaze direction and gestures were translated into visible system actions, allowing users to immediately understand the outcome of their interaction. Real-time visual feedback enhanced transparency and user confidence, ensuring that system responses were intuitive and easy to validate.

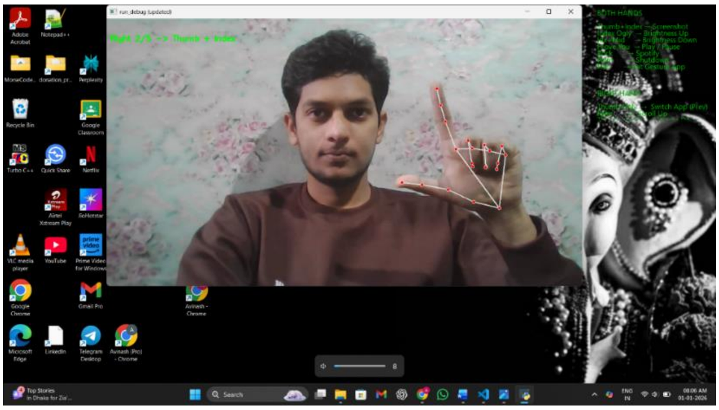

Fig. 3. Gesture System Control

Impact on System Efficiency

- **Low Computational Overhead:** The Air Command system operates efficiently on consumer-grade hardware with minimal computational load. The use of lightweight vision-processing pipelines enables real-time interaction without affecting overall system responsiveness.
- **Efficient Frame Processing:** Only essential visual features such as facial landmarks, eye regions, and hand positions are processed from each video frame. This selective processing reduces unnecessary computation and ensures smooth interaction even during continuous usage.
- **Local and Secure Data Handling:** All video processing and interaction logic are executed locally on the user's machine. No visual data is transmitted or stored externally, ensuring controlled data flow, user privacy, and improved system reliability.
- **Scalable and Modular Design:** The modular architecture of the system allows new gestures, interaction rules, and control features to be added without impacting existing functionality. This design supports future expansion while maintaining consistent performance.

V. RESULTS AND DISCUSSION

The experimental evaluation of the Air Command: A Vision-Driven Gesture and Gaze Control System demonstrates the effectiveness of vision-based interaction for touch-free computer control. The system showed consistent and accurate performance in interpreting gaze direction for cursor movement and hand gestures for command execution under normal operating conditions. This confirms that camera-based interaction techniques can serve as a practical alternative to traditional input devices without requiring specialised hardware.

By integrating real-time computer vision processing with system-level control mechanisms, the proposed system enables responsive interaction with minimal latency. Cursor movement was smooth and stable due to the incorporation of dead-zone handling and motion scaling, while gesture hold-time and cooldown logic reduced unintended command execution. Real-time visual feedback further improved user awareness and interaction confidence.

The evaluation also indicates that the system operates efficiently with low computational overhead, as only essential visual features are processed from each video frame. Local processing of visual data ensures user privacy and eliminates dependency on external servers. Overall, the results demonstrate that the Air Command system is reliable, user-friendly, and suitable for real-world touch-free interaction scenarios, including accessibility-focused and hygienic environments.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁝ Peer-reviewed & Refereed journal ⁝ Vol. 15, Issue 1, January 2026
DOI: 10.17148/IJARCCE.2026.15182

VI. CONCLUSION

This paper presented **Air Command**, a vision-driven gesture and gaze control system designed to enable touch-free computer interaction using a standard webcam. The system integrates gaze-based cursor control with hand gesture-based command execution, providing a natural and intuitive interaction model without reliance on physical input devices or specialised hardware.

Experimental evaluation confirmed that the system delivers stable real-time performance, accurate interaction detection, and efficient execution of system commands across diverse usage scenarios. The clear separation between cursor control and command actions, along with stability-enhancing mechanisms, improves usability and reduces cognitive load. Overall, the proposed system proves to be a practical, cost-effective, and accessible solution for touch-free human–computer interaction, contributing toward more inclusive and intuitive computing environments.

VI. FUTURE WORK

Future enhancements of the Air Command system will focus on expanding interaction capabilities and improving adaptability. Additional gestures and customisable interaction profiles can be introduced to support a wider range of applications and user preferences. Incorporating machine learning–based gesture adaptation may further improve recognition accuracy across diverse users.

The system can also be extended to support mobile and wearable platforms, enabling touch-free interaction beyond desktop environments. Integration with augmented reality and virtual reality applications presents another promising direction. Additionally, long-term usability studies and performance optimisation under varying lighting conditions will further enhance system robustness and real-world applicability.

REFERENCES

[1]. Szeliski, R., Computer Vision: Algorithms and Applications, 2nd edition, Springer, Cham, Switzerland, 2022. https://link.springer.com/book/10.1007/978-3-030-34372-9

[2]. Bulling, A., Ward, J. A., Gellersen, H., and Tröster, G., Eye movement analysis for activity recognition using electrooculography, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 4, pp. 741–753, 2011. https://ieeexplore.ieee.org/document/5714027

[3]. Mitra, S., and Acharya, T., Gesture recognition: A survey, IEEE Transactions on Systems, Man, and Cybernetics, vol. 37, no. 3, pp. 311–324, 2007. https://ieeexplore.ieee.org/document/4209704

[4]. Nielsen, J., Usability Engineering, Morgan Kaufmann Publishers, San Francisco, USA, 1994. https://www.sciencedirect.com/book/9780125184069/usability-engineering

[5]. Turk, M., and Pentland, A., Face recognition using eigenfaces, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Maui, Hawaii, USA, pp. 586–591, 1991. https://ieeexplore.ieee.org/document/139758

[6]. Hansen, D. W., and Ji, Q., In the eye of the beholder: A survey of models for eyes and gaze, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 3, pp. 478–500, 2010. https://ieeexplore.ieee.org/document/5206635

[7]. Starner, T., Weaver, J., and Pentland, A., Real-time American Sign Language recognition using desk and wearable computer based video, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 12, pp. 1371–1375, 1998. https://ieeexplore.ieee.org/document/735806

[8]. Bradski, G., and Kaehler, A., Learning OpenCV: Computer Vision with the OpenCV Library, O’Reilly Media, Sebastopol, USA, 2008. https://www.oreilly.com/library/view/learning-opencv/9780596516130/

[9]. Shneiderman, B., Designing the User Interface: Strategies for Effective Human–Computer Interaction, 5th edition, Pearson Education, Boston, USA, 2010. https://www.pearson.com/en-us/subject-catalog/p/designing-the-user-interface/P200000006728