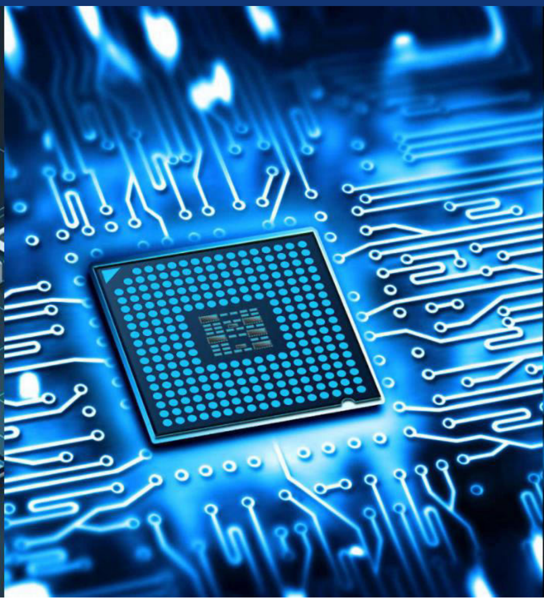
IJIRCCE

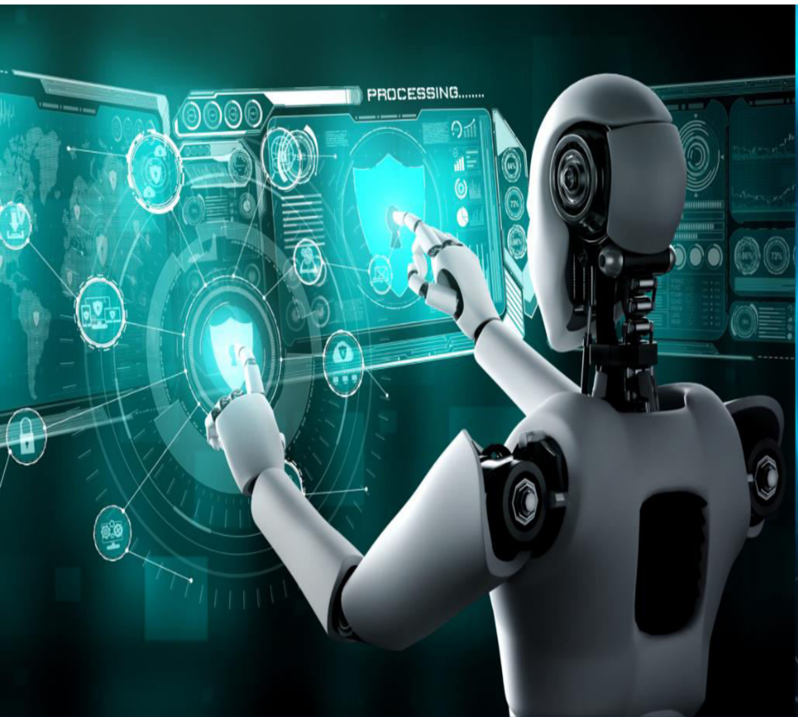
ISSN(O): 2320-9801

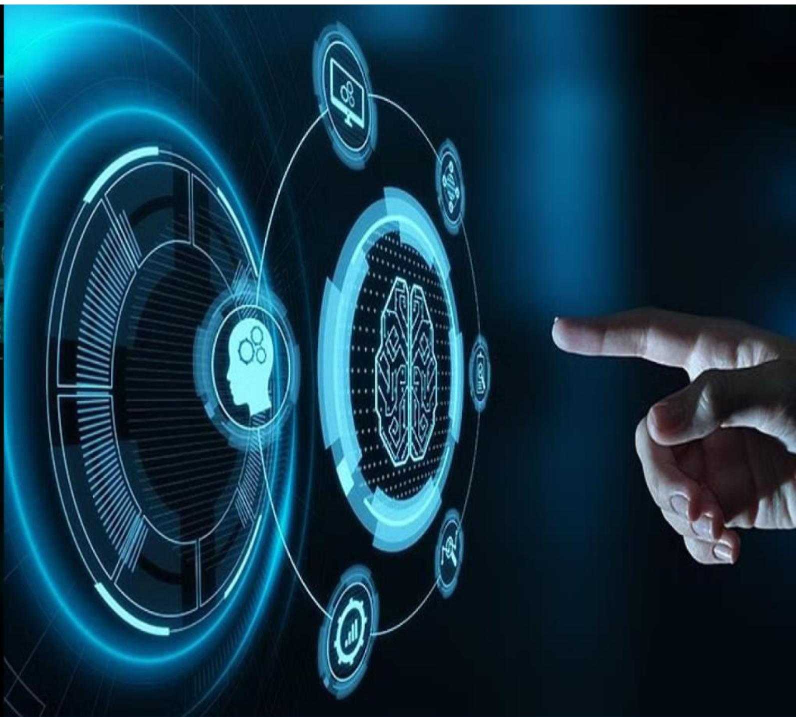
ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 8.771

Volume 14, Issue 3, March 2026

© 2026 IJIRCCE | Volume 14, Issue 3, March 2026 DOI: 10.15680/IJIRCCE.2026.1403060

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Gesture Cam- Controlled Mouse System with Voice Control and Hand Gestures

G.V.S Ananthanath[1], P.Kannathalli[2], K.Adhi Karthik[3], Y.Manikanta[4], T.Harish[5], G.Anil[6]

Sr Professor, Dept. of CSE(AI&ML), NSRIT, Visakhapatnam, AP, India[1]

Dept. of CSE(AI&ML), NSRIT, Visakhapatnam, AP, India[2,3,4,5,6]

ABSTRACT: Traditional input devices such as the mouse and keyboard limit natural interaction between humans and computers. With advances in computer vision and speech recognition, gesture- and voice-based interfaces offer a more intuitive and contactless alternative. This paper presents a real-time hand gesture and voice-controlled Human–Computer Interaction (HCI) system using MediaPipe, OpenCV, and PyAutoGUI. The proposed system enables cursor movement, mouse clicks, scrolling, window control, application launching, PowerPoint presentation control, and air-drawing through dynamic hand gestures captured via a standard webcam. Additionally, a multi-threaded voice command module allows hands-free execution of system-level operations. Experimental results demonstrate high responsiveness, accuracy, and usability under normal lighting conditions, making the system suitable for accessibility support, touchless computing, and smart environments.

KEYWORDS: Hand Gesture Recognition, Human–Computer Interaction, MediaPipe, Computer Vision, Voice Control, Touchless Interface

I. INTRODUCTION

Human–Computer Interaction (HCI) has evolved significantly from early command-line interfaces (CLI) to graphical user interfaces (GUI) and, more recently, to natural user interfaces (NUI) that aim to replicate intuitive human behavior. While traditional input devices such as keyboards and mice are efficient, they require physical contact and fine motor control, which may not be suitable for all users or environments. As computing systems become more pervasive, there is an increasing demand for interaction techniques that are natural, touchless, and accessible.

Gesture-based interaction enables users to communicate with digital systems using natural hand movements, closely mimicking real-world human communication. Such interaction paradigms are particularly beneficial for individuals with physical disabilities, elderly users, and scenarios where physical contact is undesirable or impractical. Applications such as public information kiosks, healthcare systems, smart classrooms, and virtual or hybrid presentations greatly benefit from touch-free interaction, especially in hygienic or shared environments.

Recent advances in computer vision and machine learning have made it possible to achieve accurate hand tracking using standard RGB cameras without the need for specialized hardware. Frameworks such as MediaPipe provide real-time hand landmark detection with high precision and low computational cost, enabling practical deployment on consumer-grade systems. These developments have opened new opportunities for low-cost, vision-based HCI systems that are both scalable and user-friendly.

This research proposes a robust real-time gesture-controlled mouse system enhanced with integrated voice control to provide a multimodal HCI solution. Unlike traditional approaches that rely on external sensors, depth cameras, or data gloves, the proposed system uses only a standard webcam and software-based hand landmark detection. By combining vision-based gesture recognition with speech-based commands, the system allows users to interact with the operating system in a flexible and natural manner.

The proposed system supports advanced interaction features such as adaptive cursor smoothing based on hand velocity, gesture stability checking to reduce false activations, PowerPoint presentation control through hand swipes, and an air-drawing mode using pinch gestures. These features extend beyond basic cursor control and demonstrate the system's applicability in real-world scenarios such as presentations, creative tasks, and assistive computing. Overall, the work

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

aims to contribute toward the development of practical, affordable, and intuitive next-generation Human–Computer Interaction systems.

II. LITERATURE REVIEW

Human–Computer Interaction (HCI) has evolved toward more natural and intuitive interfaces, with hand gesture recognition emerging as a promising alternative to traditional input devices such as the mouse and keyboard. Gesture-based interaction enables users to control computer systems using natural hand movements, reducing dependency on physical hardware and improving accessibility. A detailed survey of hand gesture recognition techniques based on computer vision is presented in [1]. This work compares vision-based and sensor-based methods across multiple parameters such as datasets, detection range, segmentation techniques, and classification approaches. The study emphasizes that vision-based systems using cameras are more cost-effective, flexible, and user-friendly, as they eliminate the need for external devices.

With the emergence of deep learning techniques, static hand gesture recognition has achieved improved accuracy and robustness. In [2], a Convolutional Neural Network (CNN)–based approach was proposed to handle variations in hand orientation, shape, and size among different users. Data augmentation techniques such as rotation, scaling, and shifting were applied to enhance classification performance. While effective, CNN-based systems often require large datasets and higher computational resources, which can limit real-time usability on standard hardware.
Several researchers have focused on using hand gestures for real-time computer control.

The system described in [3] enables mouse operations, media control, and shortcut execution using static hand gestures. Principal Component Analysis (PCA) was used for gesture recognition; however, such traditional techniques struggle with gesture stability and accuracy in dynamic environments. Similarly, a camera-based hand gesture recognition system using motion history images was proposed in [4] to classify dynamic gestures efficiently. Although computationally simple, the approach requires controlled lighting and background conditions, reducing robustness in real-world scenarios.

In addition to gesture-based control, voice interaction has been explored to enhance system usability and automation. A voice-controlled personal assistant device was presented in [5], enabling users to perform tasks through speech commands. Voice assistants provide hands-free interaction and improve accessibility, especially when combined with other interaction modalities. A hybrid gesture-controlled mouse and voice assistant system was proposed in [6], where cursor movement is controlled using real-time camera input while voice commands assist in task management. However, reliance on color-based hand detection limited performance under varying lighting conditions.

Further studies such as [7] and [8] explored hand gesture recognition technologies and smart systems integrating both hand gestures and voice input, highlighting the importance of real-time performance and ease of use. An OpenCV-based hand gesture mouse system with voice activation was implemented in [9], demonstrating the feasibility of webcam-based interaction. Additionally, a Python-based voice assistant system using artificial intelligence techniques was discussed in [10], focusing on user assistance and automation.
From the literature reviewed, it is observed that although gesture-based and voice-assisted

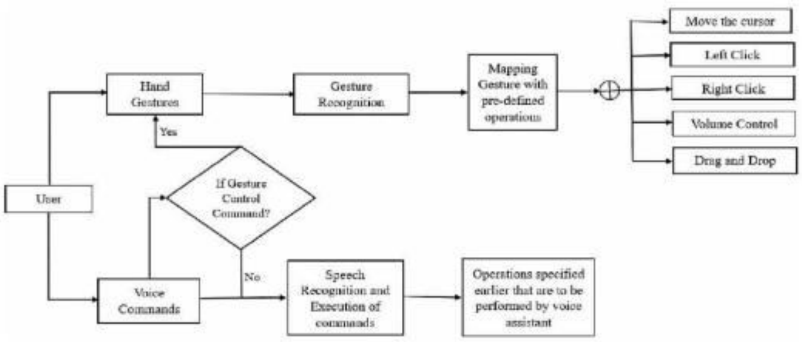
HCI systems have been widely studied, many existing solutions suffer from limitations such as dependency on controlled environments, lack of gesture stability, insufficient smoothing of cursor movement, or reliance on external hardware. Therefore, there is a clear need for a robust, real-time gesture-controlled mouse system enhanced with voice control that operates using only a standard webcam. The proposed system aims to address these limitations by incorporating adaptive smoothing, gesture stability checking, multimodal interaction, and additional functionalities such as presentation control and air drawing, making it a comprehensive and practical HCI solution.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

III. SYSTEM ARCHITECTURE



The system architecture of the proposed Gesture-Controlled Mouse with Voice Control is designed to provide a natural and hardware-free Human–Computer Interaction (HCI) experience using a standard webcam and microphone. The architecture follows a modular approach consisting of gesture input, voice input, recognition modules, decision logic, and action execution.

The system accepts input from the user in two forms: hand gestures and voice commands. The webcam continuously captures real-time video frames, which are processed to detect and track hand landmarks. These landmarks are analyzed by the gesture recognition module to identify predefined gestures. Once a gesture is recognized, it is mapped to specific mouse operations such as cursor movement, left click, right click, drag and drop, and volume control.

A decision block determines whether the detected input corresponds to a gesture-based command. If a valid gesture command is detected, the system directly executes the associated mouse operation. If no gesture command is identified, the system switches to the voice control module. Voice commands are captured through a microphone, processed using speech recognition techniques, and mapped to predefined actions such as application control or system operations.

This dual-mode interaction ensures flexibility, accessibility, and robustness. The integration of gesture and voice control reduces dependency on physical input devices and improves usability in scenarios such as presentations, touch-free environments, and assistive technologies.

IV. METHODOLOGY

The proposed system implements a real-time gesture-controlled mouse integrated with voice control to provide a natural and contactless Human–Computer Interaction (HCI) experience. The methodology is designed to operate using only a standard webcam and microphone, eliminating the need for external sensors or hardware devices. The complete workflow of the system is divided into the following stages.

4.1 Video Capture
A standard webcam is used to continuously capture real-time video frames of the user’s hand. The video stream is processed frame by frame to ensure low latency and smooth interaction. Each frame is resized and converted into a suitable color format to improve processing efficiency.

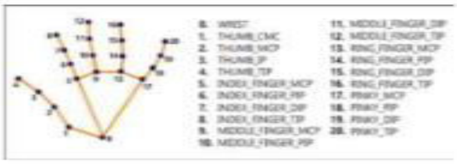
4.2 Hand Detection and Landmark Extraction
The captured frames are processed using a computer vision–based hand detection model. Once a hand is detected, a landmark extraction technique is applied to identify 21 key points on the hand, including the wrist, finger joints, and fingertips. These landmarks represent the spatial structure of the hand and are essential for understanding hand posture and motion.
The landmark-based approach allows accurate tracking of finger movement and hand orientation in real time, even under moderate variations in lighting and background.

© 2026 IJIRCCE | Volume 14, Issue 3, March 2026 DOI: 10.15680/IJIRCCE.2026.1403060

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

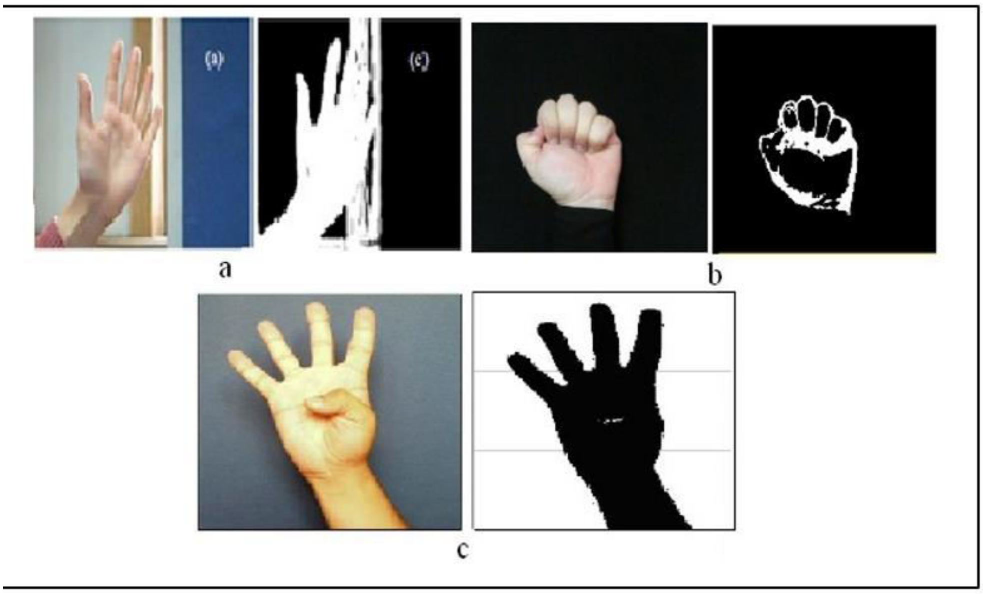
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



4.3 Gesture Recognition

Gesture recognition is performed by analyzing the relative positions and distances between the extracted hand landmarks. Specific gestures are defined based on finger states (extended or folded) and inter-landmark relationships. Dynamic gestures are recognized by observing landmark movement across consecutive frames.

To improve accuracy and prevent false detections, gesture stability checking is implemented. A gesture is confirmed only if it is detected consistently for a predefined number of frames. Additionally, adaptive smoothing techniques are applied to reduce cursor jitter and ensure smooth pointer movement.



4.4 Gesture-to-Action Mapping

Once a gesture is successfully recognized, it is mapped to a corresponding mouse operation. The predefined gesture-action mappings include:

- Cursor movement using index finger motion
- Left click using a pinch gesture
- Right click using a specific finger combination
- Drag and drop using sustained gesture detection
- Volume control using vertical hand movement

These actions are executed using system automation libraries, enabling direct interaction with the operating system.

4.5 Decision Logic for Multimodal Interaction

The system includes a decision module that determines whether the user input corresponds to a gesture-based command or a voice-based command. If a valid gesture command is detected, the system prioritizes gesture execution. If no gesture is detected, the system activates the voice recognition module. This decision logic ensures seamless switching between gesture and voice interaction modes.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

4.6 Voice Command Processing
Voice commands are captured through a microphone and processed using speech-to-text techniques. The recognized text is compared with a predefined set of voice commands. Once a match is found, the corresponding operation is executed. Voice commands are used for tasks such as application control, system navigation, and auxiliary functions, enhancing hands-free interaction.

4.7 Command Execution
The final stage involves executing either gesture-based mouse actions or voice-based system commands in real time. The system continuously loops through all the above stages to provide uninterrupted and responsive interaction. Proper error handling is implemented to avoid unintended operations.

4.8 Overall System Flow
The system continuously captures user input, processes it through gesture recognition and voice recognition modules, and executes the corresponding actions. This multimodal approach improves usability, accessibility, and efficiency, making the system suitable for real-world applications such as virtual presentations, touch-free environments, and assistive technologies.

V. IMPLEMENTATION DETAILS

The proposed Real-Time Gesture-Controlled Mouse System with Voice Control is implemented using Python, chosen for its simplicity, extensive library support, and suitability for rapid prototyping of computer vision and automation applications. The system is designed to run efficiently on a Windows operating system using commonly available hardware.

5.1 Programming Language
- Python is used as the primary programming language due to its flexibility and strong ecosystem for image processing, machine learning, and system automation.

5.2 Software Libraries Used
- OpenCV: Used for real-time video capture, frame preprocessing, and basic image operations. OpenCV ensures efficient handling of webcam input and supports real-time performance.
- MediaPipe: Utilized for accurate hand detection and landmark extraction. MediaPipe provides a robust hand tracking model that detects 21 key landmarks on the hand, enabling precise gesture recognition without external sensors.
- NumPy: Used for numerical computations, array manipulation, and mathematical operations on hand landmark coordinates.
- PyAutoGUI: Responsible for executing mouse-related operations such as cursor movement, clicking, dragging, and scrolling based on recognized gestures.
- SpeechRecognition:
Used to capture voice input and convert speech into text. Recognized commands are matched with predefined instructions to perform system-level actions.
- PyGetWindow: Used for window management tasks such as detecting active windows and controlling presentation applications like PowerPoint.

5.3 Hardware Requirements
- Standard Webcam: For capturing real-time hand gestures
- Microphone: For receiving voice commands
- No external sensors or wearable devices are required

5.4 Operating System
- The system is developed and tested on Windows OS, ensuring compatibility with commonly used desktop environments and presentation tools.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

5.5 Performance Optimization

To achieve real-time performance, the system incorporates several optimization strategies:

- Minimization of frame buffering to reduce latency
- Efficient data structures for landmark processing
- Adaptive smoothing techniques to stabilize cursor movement
- Gesture stability verification to prevent accidental actions

These optimizations allow the system to operate smoothly even on moderate hardware configurations.

VI. EXPERIMENTAL RESULTS

The proposed system was evaluated through extensive testing under different environmental and usage conditions to assess its performance, accuracy, and usability.

6.1 Testing Conditions

- Various lighting conditions (normal indoor lighting, bright light, low light)
- Different hand movement speeds
- Multiple users with varying hand sizes and gestures

6.2 Observations and Performance Metrics

- **Average Frame Rate:** 25–35 FPS, ensuring smooth real-time interaction
- **Gesture Accuracy:** High accuracy observed for cursor movement, left click, and right click gestures
- **Latency:** Minimal delay in voice command execution
- **Stability:** PowerPoint slide navigation performed reliably with very low false positives

6.3 User Feedback

User testing revealed improved usability compared to traditional mouse-only interaction, particularly during presentations. Users reported reduced physical strain and enhanced convenience when switching between gesture and voice control.

VII. APPLICATIONS

The proposed system has a wide range of real-world applications, including:

- **Touchless Computer Interaction:** Useful in environments where physical contact with devices should be minimized
- **Assistive Technology:** Enables physically challenged users to interact with computers more easily
- **Smart Classrooms and Presentations:** Allows presenters to control slides and cursor movement remotely
- **Virtual Drawing and Creative Applications:** Enables air drawing and gesture-based creativity
- **Public Kiosks and Hygienic Environments:** Suitable for hospitals, public terminals, and shared systems

VIII. LIMITATIONS

Despite its effectiveness, the system has certain limitations:

- Performance degrades under poor or inconsistent lighting conditions
- Initial user calibration is required for optimal accuracy
- Background clutter may affect hand detection accuracy
- Voice recognition performance depends on internet connectivity and ambient noise

IX. FUTURE WORK

Future enhancements can further improve the system’s robustness and functionality:

- Implementation of **deep learning-based gesture classification** for higher accuracy
- Support for **custom gesture training** by users

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

- **Multi-user support** for collaborative environments
- **Cross-platform compatibility** for Linux and macOS
- Integration with **AR/VR systems** for immersive interaction experiences

X. CONCLUSION

This project successfully presents a real-time gesture-controlled mouse system enhanced with voice control, offering an intuitive and touchless approach to Human–Computer Interaction. By utilizing only a standard webcam and microphone, the proposed system eliminates the need for external hardware, making it cost-effective, portable, and easy to deploy.

The system effectively combines computer vision–based hand gesture recognition with speech recognition to enable seamless control of cursor movement, mouse clicks, drag-and-drop operations, volume control, and presentation navigation. The use of hand landmark detection and gesture stability mechanisms ensures accurate and smooth interaction, while adaptive smoothing minimizes cursor jitter. Experimental results demonstrate that the system operates at a real-time frame rate with high accuracy and minimal latency, even under varying usage conditions.

User testing confirms that the proposed solution improves usability compared to traditional mouse-based interaction, particularly in presentation scenarios and touch-free environments. Although the system has certain limitations related to lighting conditions and background complexity, its overall performance validates the feasibility of multimodal interaction using vision and voice.

In conclusion, the proposed gesture-controlled mouse with voice assistance represents a practical and efficient HCI solution with significant potential for real-world applications. With further enhancements such as deep learning-based gesture recognition and cross-platform support, the system can be extended to serve a wider range of users and advanced interactive environments.

REFERENCES

[1] M. Oudah, “Hand Gesture Recognition Based on Computer Vision,” Journal of Imaging, Middle Technical University, Baghdad, Iraq, Jul. 23, 2020.
[2] Md. Z. Islam, “Static Hand Gesture Recognition using Convolutional Neural Network,” ResearchGate, Dept. of Computer Science and Engineering, University of Chittagong, Bangladesh, Apr. 2019.
[3] M. V. Gore, “Human Computer Interaction using Hand Gesture Recognition,” International Journal of Engineering Research & Technology (IJERT), Govt. College of Engineering, Aurangabad, India, Apr. 2014.
[4] V. Shinde, T. Bacchav, J. Pawar, and M. Sanap, “Hand Gesture Recognition System Using Camera,” International Journal of Engineering Research & Technology (IJERT), Navsahyadri Education Society’s Group of Institutions, Pune, India, Jan. 2014.
[5] A. Dekate, C. Kulkarni, and R. Killedar, “Study of Voice Controlled Personal Assistant Device,” International Journal of Computer Trends and Technology (IJCTT), AISSMS College of Engineering, Pune, India, Dec. 2016.
[6] L. Gagnani, H. Patel, S. Chaturvedi, and R. Jaiswal, “Gesture Controlled Mouse and Voice Assistant,” International Journal for Research in Applied Science & Engineering Technology (IJRASET), LDRP Institute of Technology and Research, Gandhinagar, Gujarat, India, Oct. 10, 2022.
[7] R. P. Kshirsagar, “Hand Based Gesture Recognition Technologies,” Government Polytechnic Aurangabad, Maharashtra, India, Jun. 2019.
[8] G. B. Girish, “A Smart System using Hand Gestures and Voice,” Dept. of Computer Science and Engineering, Jun. 2022.
[9] S. Raj N, “Implementing Hand Gesture Mouse Using OpenCV,” International Journal of Engineering Research & Technology (IJERT), Guru Nanak Dev Engineering College, Bidar, Karnataka, India, Jun. 2020.
[10] D. Pandey, “Voice Assistant Using Python and AI,” Dept. of Computer Science and Engineering, Rameshwaram Institute of Technology & Management, Lucknow, India, May 2022.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING