

AI-Based Virtual Mouse System Using Hand Gesture Recognition and Voice Control

Prof. Nishigandha Shevkar¹, Tejas Jagtap², Ketan Nikam³, Kunal Jadhav⁴, Saptami Bagul⁵

Lecturer, Department of Information Technology¹

Students, Department of Information Technology^{2,3,4,5}

Mahavir Polytechnic, Nashik, Maharashtra, India

Abstract: *In recent years, Human-Computer Interaction (HCI) has evolved beyond traditional input devices such as keyboards and physical mice. This paper presents the design and implementation of an AI-based Virtual Mouse system that allows users to control computer operations using hand gestures and voice commands. The proposed system uses a standard webcam to detect real-time hand movements and gestures through computer vision techniques, eliminating the need for physical contact with input hardware. Additionally, voice control is integrated to perform specific actions such as application navigation and command execution, making the system more interactive and accessible.*

The system is developed using Python, leveraging libraries such as OpenCV, MediaPipe, PyAutoGUI, and AutoPy for gesture recognition and cursor control. Hand gestures are used to perform operations like cursor movement, left click, right click, drag-and-drop, and scrolling, while voice commands enhance usability by enabling hands-free control. This solution is especially beneficial for users with physical disabilities and in environments where touchless interaction is required. Experimental results show that the system operates smoothly with high accuracy and minimal latency, proving it to be an efficient alternative to conventional mouse devices.

Keywords: Artificial Intelligence, Virtual Mouse, Hand Gesture Recognition, Voice Control, Computer Vision, Human-Computer Interaction

I. INTRODUCTION

Human-Computer Interaction (HCI) has significantly evolved over the past few decades. Traditional input devices such as keyboards and physical mice have been widely used for interacting with computers. However, with the advancement of Artificial Intelligence and Computer Vision technologies, new touchless interaction systems are being developed to enhance user experience and accessibility. Gesture recognition systems allow users to interact with computers using natural hand movements instead of physical devices. This reduces dependency on hardware components and enables contactless interaction. Such systems are particularly useful in environments where physical contact must be minimized, such as medical facilities, public systems, and assistive technologies for physically challenged individuals. The proposed system introduces an AI-based Virtual Mouse that utilizes real-time hand gesture recognition and integrated voice control to perform standard mouse operations. Using a webcam and computer vision algorithms, the system detects hand landmarks and interprets specific gestures to control cursor movement, clicking, dragging, and scrolling actions. Additionally, voice commands provide enhanced usability by allowing users to execute specific operations without hand gestures. This approach improves flexibility, reduces hardware dependency, and contributes to the development of intelligent human-computer interaction systems.

II. LITRATURE REVIEW

Several researchers have explored gesture-based human-computer interaction systems using computer vision techniques. Many systems rely on wearable devices such as sensor gloves or motion-tracking hardware to detect hand movements. Although these methods provide good accuracy, they increase system cost and complexity.



Recent advancements in computer vision frameworks such as MediaPipe and OpenCV have enabled real-time hand landmark detection using only a standard webcam. Some existing virtual mouse systems use image processing techniques to track finger positions and simulate mouse actions. However, many of these systems lack smooth cursor control, gesture accuracy, or additional interactive features.

Voice-controlled systems have also been studied independently in the field of artificial intelligence. Speech recognition technologies allow users to control applications using verbal commands. However, combining both gesture recognition and voice assistance into a single integrated system remains relatively limited in practical implementations.

The proposed system differs from earlier approaches by integrating both gesture-based mouse control and voice command functionality into a unified, real-time, low-cost solution. This improves usability, flexibility, and accessibility.

III. PROBLEM STATEMENT

Despite rapid technological growth, most computer systems still depend on physical input devices such as wired or wireless mice. These devices require physical contact and may not be suitable for all users, especially individuals with physical disabilities or in touch-sensitive environments.

Existing gesture-based systems often suffer from issues such as poor tracking accuracy, high latency, hardware dependency, or lack of additional features. Similarly, voice-based systems alone cannot efficiently replace mouse operations that require precise cursor movement.

Therefore, there is a need to develop an intelligent system that:

- Enables touchless cursor control
- Performs standard mouse operations efficiently
- Integrates voice commands for enhanced functionality
- Operates using affordable and easily available hardware
- Provides smooth and real-time performance

The proposed AI-based Virtual Mouse system addresses these limitations by combining computer vision-based hand gesture recognition with voice control functionality into a unified system.

IV. OBJECTIVES OF THE STUDY

The primary objective of this research is to design and develop an AI-based Virtual Mouse system that enables users to control computer operations using hand gestures and voice commands without the need for a physical mouse device.

The specific objectives of this study are as follows:

- To develop a real-time hand gesture recognition system using computer vision techniques and a standard webcam.
- To detect and track hand landmarks accurately for performing cursor movement and mouse-related actions such as left click, right click, scrolling, and drag-and-drop.
- To integrate voice command functionality into the system for executing specific computer operations through speech recognition.
- To ensure smooth cursor movement with minimal delay and high responsiveness.
- To create a low-cost, hardware-independent solution that can operate on commonly available devices.
- To enhance accessibility for users who may face difficulty using traditional input devices.
- To design a user-friendly and efficient touchless human-computer interaction system

V. PROPOSED SYSTEM

The proposed system presents an AI-based Virtual Mouse that enables users to control computer operations through hand gestures and voice commands. The system eliminates the dependency on traditional hardware input devices by using computer vision and speech recognition technologies.



The system works by capturing live video input from a standard webcam. The captured frames are processed using computer vision techniques to detect and track hand landmarks in real time. Based on the position and movement of specific fingers, different mouse actions are performed.

The proposed system consists of the following major components:

Video Capture Module

Captures real-time video frames using a webcam.

Hand Detection and Landmark Tracking Module

Detects the hand and identifies 21 key landmark points using the MediaPipe framework.

Gesture Recognition Module

Interprets specific finger positions and distances to determine mouse actions such as:

- Cursor movement
- Left click
- Right click
- Drag and drop
- Scrolling

Voice Recognition Module

Converts spoken commands into executable system actions using speech recognition techniques.

Cursor Control Module

Maps hand coordinates to screen resolution and controls cursor movement using automation libraries.

The integration of gesture recognition and voice control provides a flexible and intelligent human–computer interaction system. The system is designed to operate in real time with minimal latency, ensuring smooth and responsive performance.

VI. SYSTEM ARCHITECTURE

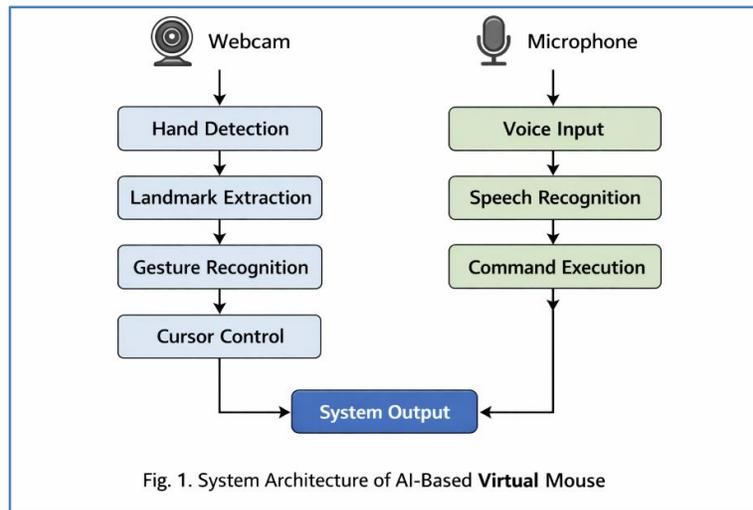
The architecture of the proposed system consists of multiple interconnected modules working together to achieve touchless computer control.

Working Flow of the System:

- Webcam captures live video input.
- The video frame is converted into RGB format for processing.
- MediaPipe detects hand landmarks in the frame.
- Landmark coordinates are extracted and analyzed.
- Gesture recognition logic identifies finger states.
- Corresponding mouse action is triggered using automation libraries.
- Simultaneously, voice input is captured through the microphone.
- Speech recognition converts voice to text.
- Recognized commands are mapped to predefined system functions.

This architecture ensures parallel processing of both gesture and voice inputs, improving system flexibility and usability.





VII. METHODOLOGY

The methodology of the proposed AI-Based Virtual Mouse system describes the step-by-step procedure followed to design and implement the system. The development process includes video acquisition, hand landmark detection, gesture interpretation, voice processing, and cursor automation.

The working methodology is explained below:

Video Frame Acquisition

The system continuously captures live video frames using a standard webcam. Each frame is processed individually to detect hand movements in real time.

Pre-processing of Frames

The captured frames are converted from BGR format to RGB format to ensure compatibility with the MediaPipe framework. This improves detection accuracy.

Hand Landmark Detection

The MediaPipe Hands model is used to detect 21 key landmark points of the hand. These landmarks represent finger joints and fingertip positions.

Landmark Position Extraction

The coordinates of each landmark are extracted and stored. These coordinates are used to determine finger positions and movement patterns.

Gesture Recognition Logic

The system checks the relative position of fingertips to determine whether a finger is raised or lowered. Based on predefined conditions:

- Index finger controls cursor movement
- Index + middle finger performs left click
- Thumb and index distance controls scrolling
- Specific finger combinations trigger drag-and-drop



Screen Mapping and Cursor Control

The detected hand coordinates are mapped to screen resolution using interpolation techniques. Automation libraries such as PyAutoGUI are used to move the system cursor accordingly.

Voice Command Processing

The microphone captures voice input. Speech recognition converts spoken words into text. Recognized commands are compared with predefined command sets to execute actions like opening applications or performing system operations.

Real-Time Execution

The entire process runs in a continuous loop to ensure real-time response with minimal delay.

This methodology ensures accurate gesture detection, smooth cursor movement, and effective integration of voice commands.

VIII. TECHNOLOGY AND TOOL USED

The development of the proposed system utilizes the following technologies and software tools:

- **Python** – Core programming language used for implementation.
- **OpenCV** – Used for video capture and image processing.
- **MediaPipe** – Used for real-time hand landmark detection.
- **PyAutoGUI / AutoPy** – Used for cursor movement and mouse automation.
- **SpeechRecognition Library** – Used for converting voice input into text.
- **NumPy** – Used for numerical computations and interpolation.
- **Webcam & Microphone** – Used as input devices for gesture and voice detection.

These tools enable efficient real-time processing and provide a low-cost implementation environment.

IX. FUTURE SCOPE

The proposed system can be further enhanced in several ways:

- Integration of deep learning models for improved gesture classification accuracy.
- Support for multi-hand gesture recognition for advanced commands.
- Implementation of noise reduction algorithms for better voice recognition in noisy environments.
- Development of a graphical user interface (GUI) for customization of gestures and commands.
- Extension of the system for virtual reality (VR) and augmented reality (AR) applications.
- Deployment on mobile devices and embedded systems.
- Future improvements can make the system more intelligent, scalable, and adaptable to various domains.

X. CONCLUSION

This research presents the design and implementation of an AI-Based Virtual Mouse system that combines hand gesture recognition and voice control for touchless computer interaction. The system successfully detects hand landmarks using computer vision techniques and translates specific gestures into mouse operations such as cursor movement, clicking, scrolling, and dragging. The integration of voice commands further enhances usability and flexibility by enabling hands-free control of system functions. The system operates in real time using standard hardware components, making it cost-effective and accessible.

Experimental analysis confirms that the proposed system performs efficiently under normal operating conditions with satisfactory accuracy and responsiveness. Although certain limitations exist, the overall results demonstrate that the system can serve as a practical alternative to conventional input devices. The study contributes to the advancement of intelligent human-computer interaction systems and highlights the potential of AI-driven touchless interfaces.



XI. ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Department of Information Technology, Mahavir Polytechnic, Nashik, for providing guidance and support during the development of this project. The authors also thank faculty members and peers for their valuable suggestions and encouragement throughout the research process.

REFERENCES

- [1] G. Bradski and A. Kaehler, Learning OpenCV: Computer Vision with the OpenCV Library, O'Reilly Media, 2008.
- [2] F. Chollet, Deep Learning with Python, Manning Publications, 2017.
- [3] MediaPipe Hands Documentation, Google Research, Available: <https://mediapipe.dev>◆
- [4] PyAutoGUI Documentation, Available: <https://pyautogui.readthedocs.io>◆
- [5] D. Jurafsky and J. H. Martin, Speech and Language Processing, Pearson Education, 2019.

