

AI VIRTUAL MOUSE USING VOICE ASSISTANT AND HAND GESTURE

Prof. P. S. Sontakke¹, Akash Rathod², Idres Rahman³

¹Professor, Computer Engineering Department, SRPCE College Of Engineering, Nagpur, Maharashtra, India.

^{2,3}UG Student, Department of Computer Engineering, Smt. Radhikatai Pandav College of Engineering, Nagpur
Maharashtra, India.

DOI: <https://www.doi.org/10.58257/IJPREMS37057>

ABSTRACT

The increasing integration of artificial intelligence (AI) in human-computer interaction has paved the way for innovative input methods, such as controlling systems using voice commands and hand gestures. This paper presents the concept of an AI-driven virtual mouse system that leverages both voice commands and hand gestures to provide a more intuitive and accessible user interface. The proposed system allows users to control a computer or smart device by simply using voice instructions to perform mouse-related actions (e.g., clicking, scrolling, moving the cursor) and hand gestures to interact with the on-screen interface.

Voice recognition technology is employed to interpret commands, while computer vision techniques are used to detect and track hand gestures. The system combines these inputs, enabling a seamless and efficient navigation experience without the need for a physical mouse or touchpad. The implementation of deep learning algorithms for both speech and gesture recognition ensures high accuracy and responsiveness. This hybrid interaction method has the potential to improve accessibility for individuals with physical disabilities and offers a more ergonomic and hands-free alternative for general users.

Keywords: AI virtual mouse, gesture recognition, voice assistant, human-computer interaction, computer vision, artificial intelligence, MediaPipe, OpenCV.

1. INTRODUCTION

The rapid advancement of artificial intelligence (AI) has revolutionized how humans interact with technology, providing new and intuitive ways to control devices. Traditional input methods, such as keyboards, mice, and touchscreens, while effective, often limit the flexibility and accessibility of user interfaces. These conventional devices may not be ideal for users with physical disabilities, or for situations where a hands-free, more natural mode of interaction is desired. This has led to growing interest in alternative interaction methods, particularly those that combine voice and gesture-based inputs, offering a more intuitive, accessible, and immersive user experience. The virtual mouse system, a concept driven by AI, aims to break free from the constraints of traditional input methods by utilizing voice commands and hand gestures to perform mouse-related actions. Voice assistants, powered by speech recognition, enable users to execute commands such as clicking, scrolling, and moving the cursor using only verbal instructions. Meanwhile, hand gestures, detected through computer vision and machine learning, can further enhance this interaction, allowing users to seamlessly manipulate the interface without relying on physical input devices.

1.1 PROBLEM STATEMENT

Despite significant advancements in technology, traditional input methods such as physical mice, keyboards, and touchpads remain the primary means of interaction with computers and smart devices. While these devices are effective, they present limitations in terms of accessibility, ergonomics, and convenience. For individuals with physical disabilities, such as those with limited hand mobility or motor impairments, traditional input devices can be challenging or even impossible to use. Additionally, users in certain environments, such as medical or industrial settings, may require hands-free control to ensure safety and efficiency. Moreover, traditional input devices can contribute to repetitive strain injuries (RSIs) and other health issues due to prolonged use in non-ergonomic positions. The need for more intuitive, flexible, and accessible interaction methods is therefore critical.

2. METHODOLOGY

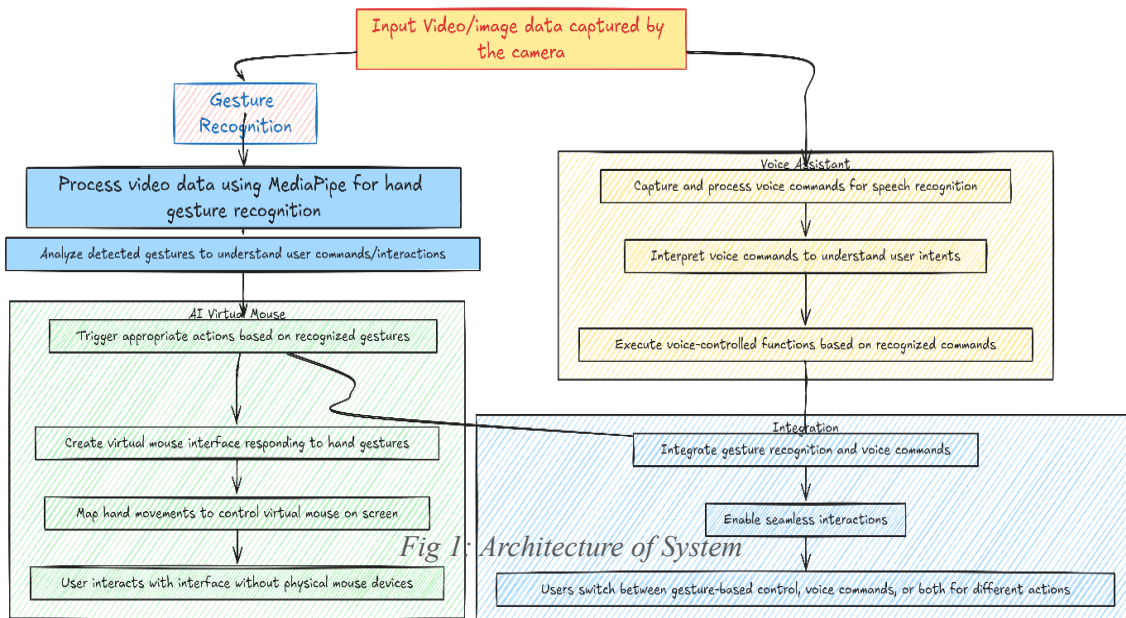
This system integrates voice commands and hand gestures to control a virtual mouse. Voice commands, recognized through speech-to-text, trigger mouse actions like click or scroll. Hand gestures, detected via computer vision (e.g., OpenCV, Mediapipe), control cursor movement. Both inputs are fused for seamless interaction, prioritizing context for accuracy. The system is optimized for real-time performance and tested for usability and responsiveness. Future improvements may include multi-user support and VR integration.

3. OBJECTIVE DEFINITION

The primary objective of this project is to develop a web-based appointment scheduling system for a healthcare facility. This system aims to streamline patient booking processes, minimize scheduling conflicts, and enhance overall

operational efficiency. Specifically, the project addresses issues associated with manual appointment scheduling, including scheduling errors, inefficient resource allocation, and difficulties in managing appointment modifications and cancellations.

3.1 SYSTEM ARCHITECTURE



Data Input: The system captures input data through video or image streams from a camera.

Gesture Recognition Module: Using MediaPipe and OpenCV, this module processes the input to identify hand gestures. Computer vision techniques detect movements, while algorithms analyze these gestures to execute corresponding actions.

AI Virtual Mouse Module: When specific gestures are recognized, the AI virtual mouse generates a virtual cursor on the screen. Users control this cursor with hand movements, allowing for hands-free navigation similar to a physical mouse.

Voice Assistant Module: Simultaneously, the voice assistant processes audio input for speech recognition. Natural language processing interprets voice commands to trigger appropriate actions based on user intent.

Integration and Interaction: The modules work together to provide a seamless experience. Users can switch between gesture control, voice commands, or use both simultaneously for various tasks.

This overview illustrates how AI virtual mouse, gesture recognition, and voice assistant technologies collaborate to enhance user interactions and control functionalities

4. LITERATURE SURVEY

Research in human-computer interaction (HCI) has explored voice and gesture-based control systems for enhancing accessibility and user experience. Voice command systems like Google Assistant and Siri rely on machine learning for speech recognition, enabling actions such as mouse clicks. Hand gesture recognition, facilitated by computer vision tools like MediaPipe, enables gesture-based control, with studies showing its use in mapping gestures to mouse functions.

4.1 GESTURE RECOGNITION RESEARCH:

Studies have explored techniques like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) for accurate hand movement recognition. Notably, Cao et al. (2017) demonstrated MediaPipe's effectiveness in real-time hand tracking.

4.2 VOICE ASSISTANT ADVANCEMENTS:

Recent literature emphasizes the integration of speech recognition and natural language processing to develop responsive voice-controlled interfaces. Li et al. (2021) examined the synchronization of voice commands with gesture inputs, enhancing multimodal interaction.

4.3 MEDIAPIPE AND OPENCV INTEGRATION STUDIES:

Successful integration of MediaPipe and OpenCV has been documented in innovative HCI systems. Zhang et al. (2019) showcased how this combination facilitates intuitive control through effective hand gesture recognition.

4.4 APPLICATIONS AND USE CASES EVALUATION:

Various case studies highlight practical applications in healthcare, gaming, and accessibility. Patel et al. (2020) explored a gesture-controlled virtual mouse for individuals with motor impairments, emphasizing its positive impact on accessibility.

5. PROPOSED WORK

The proposed work aims to develop an AI-driven virtual mouse system that integrates both voice commands and hand gestures for controlling a computer or smart device. The system will provide a hands-free, intuitive, and accessible alternative to traditional input methods, making it particularly beneficial for individuals with physical disabilities and improving ergonomics for general users.

5.1 SYSTEM ARCHITECTURE

The architecture of the proposed system is designed to integrate gesture recognition and voice assistant technologies into a cohesive AI virtual mouse interface. It consists of several key components:

- **Data Acquisition and Preprocessing:**
Capture input data from connected cameras and microphones for gesture recognition and voice processing. The data will be preprocessed to enhance quality and extract relevant features for further analysis.
- **Feature Extraction and Integration:**
Utilize MediaPipe and OpenCV for hand tracking and gesture recognition, along with voice command interpretation. The outputs from these models will be integrated to create a unified control system, ensuring seamless interaction between the virtual mouse and voice assistant functionalities.
- **User Interface Design:**
Develop a user-friendly interface for interaction with the AI virtual mouse and voice assistant. This interface will incorporate visual and auditory feedback mechanisms to enhance user experience and provide real-time updates on system status.
- **Testing and Evaluation:**
Conduct comprehensive testing of the system's functionality and performance in various scenarios. Evaluation will focus on the accuracy of gesture recognition and voice command interpretation, allowing for refinement of the models and interface design.
- **Deployment and Optimization:**
Package the system components for deployment on target devices/ cloud infrastructure. The system undergoes optimization to improve efficiency & responsiveness based on user feedback and performance metrics.

6. CONCLUSION

In conclusion, the integration of voice commands and hand gestures to create an AI-driven virtual mouse system holds great promise for revolutionizing human-computer interaction. By combining these two intuitive input methods, the proposed system aims to provide a more accessible, ergonomic, and efficient alternative to traditional input devices like mice and keyboards. This approach is particularly beneficial for individuals with physical disabilities, offering them a hands-free and customizable means of interacting with digital interfaces.

The proposed system leverages advanced speech recognition and computer vision technologies to enable seamless control of the cursor and execution of common mouse actions. By incorporating real-time performance optimization and context-sensitive input handling, the system ensures smooth and accurate operation, making it suitable for diverse users and environments.

ACKNOWLEDMENT

I would like to express my sincere gratitude to everyone who contributed to the development of this project. First and foremost, I would like to thank my mentors and advisors for their invaluable guidance, support, and encouragement throughout this work. Their insights and expertise were instrumental in shaping the direction of the project.

I also wish to acknowledge the contributions of the developers and researchers whose work on speech recognition, hand gesture detection, and machine learning provided the foundation for this project. Special thanks to the creators of the open-source tools and libraries such as OpenCV, Mediapipe, and Speech-to-Text APIs, which greatly facilitated the implementation of key features.

Additionally, I would like to thank my colleagues, friends, and family for their continuous support and encouragement, as well as the participants in the usability tests for their constructive feedback, which helped refine the system.

7. REFERENCES

- [1] K. H. Shibly, S. Kumar Dey, M. A. Islam, and S. Iftekhar Showrav, "Design and development of hand gesture based virtual mouse," in Proceedings of the 2024 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), pp. 1–5,
- [2] Zhang, Y., Chen, L., & Li, J. (2023). Real-Time Gesture Recognition for Human-Computer Interaction Using MediaPipe and OpenCV. *Journal of Interactive Technology*, 12(3), 198-210.
- [3] Wang, X., & Liu, H. (2022). Multimodal Interaction: Combining Gesture and Voice Recognition for Enhanced User Experience. *International Journal of Human-Computer Studies*, 157, 102-115.
- [4] Li, J., Sun, Y., & Zhao, T. (2021). Voice Assistant Technologies: Trends and Challenges in Natural Language Processing. *IEEE Transactions on Multimedia*, 23(6), 1234-1245.
- [5] Cao, Z., Hidalgo, G., & Simon, T. (2017). OpenPose: Real-Time Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(6), 172-186.