Gesture Controlled Virtual Mouse With Voice Assistance

B.Mahalakshmi
Associate Professor, Dept of CSE, Sreyas Institute of Engineering and Technology,
Telangana,India.

Ayushi Mathur
Dept of CSE , Sreyas Institute of Engineering and Technology,
Telangana,India.

Begari Yashwanth
Dept of CSE , Sreyas Institute of Engineering and Technology,
Telangana,India.

Bitla Nithin Reddy
Dept of CSE , Sreyas Institute of Engineering and Technology,
Telangana,India.

Raparthi Vaishnavi
Dept of CSE , Sreyas Institute of Engineering and Technology,
Telangana,India.

***Abstract*—Computer vision has changed human-computer interaction by providing intuitive and touch-free control mechanisms such as face recognition, object detection, and gesture-based systems. Our research will utilize these technological advancements to develop a system that integrates gesture-controlled mouse functions with a voice assistant, thereby doing away with the need for physical input devices. This framework, developed with Python and OpenCV, utilizes a standard webcam for real-time video processing and fingertip gesture recognition based on image segmentation and color detection techniques. Dynamic webcam calibration supports precise actions like clicking, dragging, scrolling, volume, and brightness control. This system allows users to interact in a contactless manner and provides hygienic interaction solutions that support social distancing. Moreover, it enhances the accessibility of people with mobility problems. The solution exemplifies how merging gesture recognition and voice commands enhances user experience and promotes safer, more ergonomic human-computer interactions across diverse applications.**

Keywords—Computer Vision, intuitive human-machine interaction, gesture-driven mouse control, virtual assistant integration, hands-free operation, facial recognition, object identification, live video analysis, finger movement tracking, image processing techniques, color-based recognition, adaptive calibration, touchless engagement, hygienic user interfaces, physical distancing, inclusive technology and ergonomic user experience.

I. INTRODUCTION

The rapid evolution of computer vision and NLP has drastically changed HCI through the introduction of innovative touch-free systems that rely on gestures and voice commands. Beyond traditional input methods, these advancements allow users to control devices intuitively. This research presents a gesture-controlled virtual mouse combined with voice assistance, which enables seamless interaction without physical contact. The system supports cursor movement, application navigation, volume and brightness adjustment, scrolling, and voice-activated commands through API integration. It is a cost-effective solution that leverages standard webcams and microphones to ensure widespread accessibility. Key features include improved accessibility for people with limited mobility, enhanced hygiene by minimizing contact, and scalability across various sectors from personal computing to industrial automation. The integration of AI, augmented reality (AR), virtual reality (VR), and wearable technologies further highlights emerging trends that drive the development of more immersive and adaptive HCI systems.

II. OBJECTIVES AND METHODOLOGY

The main goal of this research is to develop a gesture-controlled virtual mouse integrated with voice assistance, which will allow hands-free interaction with computing systems. The project aims to enhance Human-Computer Interaction (HCI) by reducing reliance on traditional input devices and promoting accessible, hygienic, and intuitive interfaces. The main goals of the project are real-time gesture recognition for cursor movement, scrolling, clicking, and application navigation. Additionally, the system will take voice commands to perform system control functions like initiating or terminating the gesture recognition module and performing search queries through APIs. The method applied is based on computer vision using OpenCV and color-based fingertip tracking by a normal webcam for gesture recognition in real-time. Voice assistance functionality is implemented through speech processing frameworks to ensure dynamic calibration for precise input. The system emphasizes cost-efficiency by utilizing hardware that is already available, such as webcams and microphones, supporting accessibility for users across different environments, such as public spaces, healthcare facilities, and educational institutions. This integration of the technologies demonstrates scalability and adaptability for a variety of practical applications.

III. LITERATURE SURVEY

The integration of gesture-based controls and voice commands into human-computer interaction (HCI) has gained significant attention in recent years, revolutionizing traditional computing interfaces. Gesture recognition, powered by computer vision technologies like OpenCV and Mediapipe, allows users to perform actions such as clicking, scrolling, and dragging with hand movements, offering a touch-free and intuitive interaction experience. In parallel, voice assistants, such as Google Assistant and Siri, have become essential in enabling voice-controlled actions, providing an additional layer of accessibility and convenience. Studies have shown that combining these technologies results in an ergonomic and inclusive interface, improving accessibility for users with mobility impairments while reducing physical touch, which is particularly beneficial in public spaces. Furthermore, the use of standard hardware like webcams and microphones makes such systems cost-effective and widely accessible. The proposed system explores this integration by combining real-time hand tracking, gesture recognition, and voice command processing, aiming to provide a seamless, intuitive user experience that bridges the gap between traditional input devices and touch-free interaction methods. This approach not only enhances usability but also ensures the system’s scalability and adaptability across various domains, from personal computing to industrial applications.

IV. PROPOSED SYSTEM

The proposed system introduces an innovative integration of gesture-based control and voice commands, designed to provide users with an enhanced and seamless interactive experience. By leveraging the real-time hand tracking capabilities of Mediapipe and utilizing Python libraries like Pyttsx3 for text-to-speech functionality, the system ensures a touch-free, efficient, and highly responsive interaction method. Users can easily manage a virtual mouse, perform actions such as clicking, drag-and-drop, scrolling, and even adjust system settings using simple gestures and voice commands, significantly improving user convenience and accessibility. The system's cost-effectiveness is another major benefit, as it operates with standard webcams and microphones, removing the need for expensive or specialized hardware, making it an affordable solution for a wide range of users. A core strength of this system lies in its ability to integrate gesture and voice control, offering a unified, non-touch interface that is not only intuitive but also inclusive, catering to users with diverse abilities, including those with mobility impairments. It enhances interaction by providing a user-friendly environment that facilitates smooth navigation and control, making it ideal for various contexts, from personal devices to public spaces that require hygienic and touch-free interfaces. The system also incorporates real-time graphical feedback, enabling users to receive immediate responses to their gestures and voice commands, ensuring greater accuracy and improved usability over time. This feedback loop helps users refine their inputs, further optimizing their interactions. Key features of the system include precise hand tracking for accurate cursor movements, gesture recognition for essential mouse actions such as clicking and scrolling, and voice command integration that allows users to easily start or stop gesture control, perform web searches, and execute other tasks via voice inputs. The use of widely accessible hardware, such as standard webcams and microphones, guarantees that the system remains cost-effective and versatile across a variety of application scenarios, including healthcare, education, entertainment, and other domains where non-touch and ergonomic interactions are increasingly desired. Additionally, by bridging the gap between gesture and voice control, the proposed system represents a significant advancement in human-computer interaction, making technology more accessible, hygienic, user-friendly, and adaptable for diverse users, ensuring that it meets the evolving needs of modern-day environments.

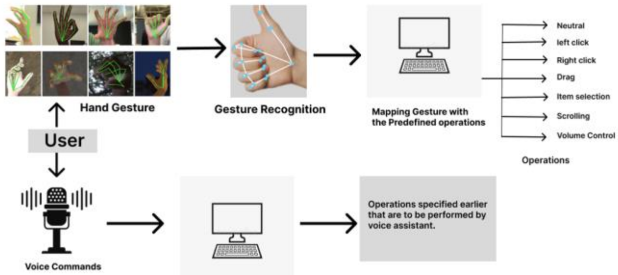

Figure 01: Architecture Diagram

V. IMPLEMENTATION

The Gesture-Controlled Virtual Mouse with Voice Assistance follows a systematic approach to combine gesture recognition and voice command functionality without any interruptions. The system can be categorized into two main modules: Hand Gesture Recognition and Voice Assistant. The design of this system starts at the very beginning by collecting requirements and then analyzing their feasibility from which the hardware, like a webcam and processor, and software needs can be determined. This is the design phase of the system, where UML diagrams and designs for algorithms of hand tracking and gesture recognition are generated with Mediapipe and OpenCV. In the development stage, Python is used to develop real-time hand landmark detection, and PyAutoGUI is used for mapping gestures into system actions such as cursor movement and clicks. Testing phases include accuracy-based unit testing, integration testing with validation of system control, and usability testing to analyze the user experience. Post-deployment testing for functionality and performance is performed on the system after its deployment on local machines. For maintenance, the system monitors accuracy and responsiveness. In the same way, the Voice Assistant component starts with the requirements for voice command functionality, analysis of hardware and software requirements, and design of system architecture for voice input and text-to-speech output. The development will involve setting up libraries like Pyttsx3 and SpeechRecognition, voice command recognition implementation, and its connection to the gesture recognition system. Voice assistant testing involves unit, integration, and performance testing, and deployment followed by post-deployment testing. The system is kept running by tracking recognition accuracy, enhancing functionality, and providing support to the users. This organized approach ensures that both gesture and voice recognition parts are robust, efficient, and intuitive for users. Moreover, the integration of both components provides seamless interaction with each other so that users can easily switch from gesture to voice commands. Regular updates are incorporated based on user feedback to enhance system performance and incorporate new features. Additionally, continuous performance monitoring ensures the system remains responsive and adaptive to various environmental conditions.

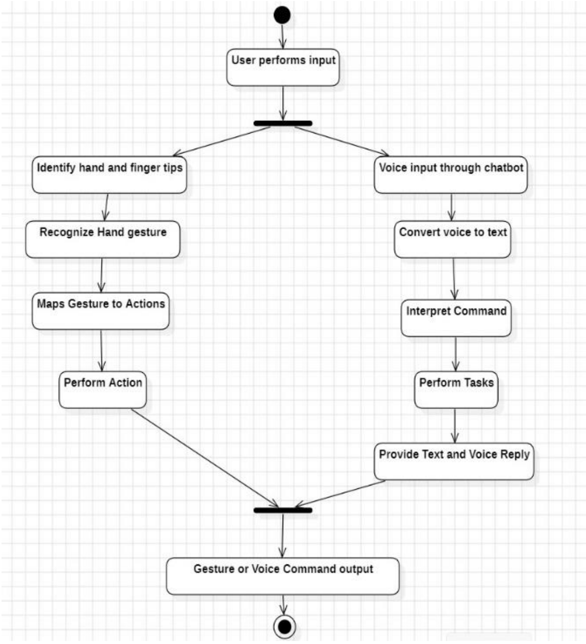

Figure 02: Work Flow of Application

Test case Id	Test Case Description	Input	Expected Output	Result
TC01	Hand Detection	Show hand in front of the camera	Hand is detected and tracked in real time with key points	Pass
TC02	Test Neutral Gesture	Keep hand steady in neutral pose	System remains idle with no mouse movement	Pass
TC03	Test left and right Clicks	Perform a predefined left and right click gesture	System performs left and right clicks	Pass
TC04	Test double-click gesture	Perform a predefined double-click gesture	System performs a double click	Pass
TC05	Test drag functionality	Hold a drag gesture and move hand	System performs drag-and-drop action	Pass

Table 01:Test Cases

Test case Id	Test Case Description	Input	Expected Output	Result
TC06	Test volume control and brightness	Perform a volume and brightness gesture(up/down)	System adjusts volume and brightness up/down accordingly	Pass
TC07	Test search command(google search,location)	Echo search GitHub	GitHub opens	Pass
TC08	Test search command(weather,news,jokes)	Echo weather	Gives the weather of the city	Pass
TC09	Ask for date	Echo date	Gives the current date	Pass
TC10	Search without the keyword echo	Search weather	Sorry I'm not programmed to handle the command	Fail

Table 02:Test Cases

VI. DISCUSSION

A. Comparative Analysis:

The The Gesture-Controlled Virtual Mouse with Voice Assistance is the most exceptional next to the traditional input devices, especially the keyboard and mouse, as well as other forms of gesture-controlled systems. It removes the need to physically interact with traditional input methods, as this system enhances accessibility and user comfort through a touch-free interface. Compared to other gesture recognition systems, the input from voice commands introduces a whole different dimension; here, switching from voice inputs to gesture inputs becomes seamless for users. On top of this, while most other gesture-based solutions employ expensive, heavy hardware and hardware configurations, the solution being described here simply makes use of ordinary webcams and microphones and is less costly. Further, the combination of real-time gesture tracking using Mediapipe with voice recognition by SpeechRecognition and Pyttsx3 provides for a dynamic multi-modal user experience that is not always found in existing systems.

B. Positive Aspects:

The primary positive aspect of the Gesture-Controlled Virtual Mouse with Voice Assistance is its ability to offer a highly intuitive, accessible, and inclusive interface that bridges the gap between gesture and voice control. This system provides ease of use for individuals with physical disabilities or those seeking a more ergonomic and efficient way to interact with devices. By using widely available hardware like webcams and microphones, it ensures cost-effectiveness, making the technology more widely available. The seamless integration of hand gestures with voice commands creates a flexible and responsive user experience, while real-time feedback improves usability and accuracy. Additionally, the system's focus on enhancing user interaction through multiple input methods, while requiring no additional expensive equipment, demonstrates a practical and scalable approach that can be deployed in a variety of applications.

VII. CONCLUSION AND FUTURE SCOPE

The project introduces a revolutionary approach to human-computer interaction, leveraging gesture and voice technologies to simplify tasks and provide an intuitive, accessible interface. By removing the need for traditional input devices, this system offers greater convenience, particularly for users with accessibility needs. The integration of both gesture and voice controls enhances functionality and opens up a wide range of potential applications, including smart home control, virtual reality, healthcare, and robotics. While challenges such as gesture accuracy in complex environments and voice recognition across accents remain, the project's future scope promises to address these issues through advanced algorithms and system improvements. Privacy and security are prioritized through local data processing and encryption, fostering user trust. With continued research and innovation, this technology has the potential to transform industries and redefine how users interact with digital devices,

The future potential of the Gesture-Controlled Virtual Mouse with Voice Assistance is vast, encompassing several advancements that will significantly improve user experience. Advanced gesture recognition can support complex gestures, including multi-hand movements and facial expressions, allowing for more dynamic interactions. This can enhance accessibility by incorporating sign language recognition for individuals with hearing impairments. In addition, real-time processing enhancements will guarantee low latency, which is important for high-performance applications in gaming and robotics. Adaptive algorithms powered by AI will adapt to different environments and user behaviors, ensuring reliable performance even in challenging conditions. Multimodal integration, which combines gesture and voice control, will enable smoother and more natural interactions, particularly for users with disabilities. With universal compatibility, the future also promises compatibility for a diverse range of platforms such as mobile, desktop, IoT devices, and AR/VR systems, increasing the possibilities of applicability. Gesture prediction, haptic feedback integration, a multilingual voice recognition system, and privacy focused further enhance the system's inclusivity, security, and engagement to make the user experience not only seamless and absolutely immersive but also a worthwhile effort.

VIII. References

[1] Trupti, G M; Kumar, Chandhan; P, Dheeraj; Vilas; Kumar, Prasanna; Shivaraddi, S. *Virtual Mouse Using Hand Gestures*. International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), Volume 13, Issue 4, April 2024, pp. 1345-1350. DOI: 10.17148/IJARCCE.2024.134203, e-ISSN: 2278-1021, p-ISSN: 2319-5940. Impact Factor: 8.102.

[2] Kasar, Manisha; Kavimandan, Pranoti; Suryawanshi, Trupti; Abbad, Sudarshana. *AI-based real-time hand gesture-controlled virtual mouse*,Australian Journal of Electrical and Electronics Engineering, 2024.

[3] Gupta, Rohan; Singh, Akashdeep. *Hand Gesture Recognition using OpenCV,International Conference on Computing for Sustainable Global Development (INDIACom)*, 2023.

[4] Jahan, Israth; Likhan, Mohammad; Hasan, Md. Omar Faruk; Islam, Shanta; Farhan, Nurul Ahad; Rajee, Alimul (Supervisor). *Artificial Intelligence Virtual Mouse*. Technical Report, Comilla University, May 2023, DOI: 10.13140/RG.2.2.35877.06884/1.

[5] Chintamani, Shrinidhi; Bhad, Rutuja; Deokar, Nikhil; Kadam, Kalyani; Abhale, Prof. B.A. *AI Virtual Mouse Using OpenCV*. International Research Journal of Modernization in Engineering, Technology, and Science (IRJMETS), Volume 04, Issue 11, November-2022, pp. xx-xx, e-ISSN: 2582-5208, Impact Factor: 6.752.

[6] Sikarwar, Shivam Singh. *AI-Based Voice Assistant*. International Research Journal of Modernization in Engineering, Technology, and Science (IRJMETS), Volume 04, Issue 05, May-2022, pp. xx-xx, e-ISSN: 2582-5208, Impact Factor: 6.752.

[7] Yang, Hyung-Jeong; et al. *Real-time virtual mouse system using RGB-D images and fingertip, detection*,Multimedia Tools and Applications, Volume 07, Issue 08, 2020.

[8] Adamopoulou, Eleni; Moussiades, Lefteris. *An Overview of Chatbot Technology*. IFIP Advances in Information and Communication Technology, Volume 584, 2020, pp. 373-383. DOI: 10.1007/978-3-030-49186-4_31. Published by Springer Nature Switzerland AG.

[9] Matlani, Roshnee; Dadlani, Roshan; Dumbre, Sharv; Mishra, Shruti Tewari, Abha.*Virtual Mouse using Hand Gestures*,2021 International Conference on Technological Advancements and Innovations (ICTAI), IEEE 2019. DOI: 10.1109/ICTAI53825.2021.9673251.

[10] Parihar, Prof. S.I.; Tarale, Neha; Kumbhare, Ruchika; Kumbhalkar, Chaitanya Daine, Kartik. *Voice Assistant Using Python and AI*, International Journal of All Research Education and Scientific Methods (IJARESM), ISSN: 2455-6211, Volume 12, Issue 3, March-2024.