International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 4,2025
https://theaspd.com/index.php

Hands-Free Mouse With Gesture And Voice Control

Manmath Nath Das[1] ,Jyotsnarani Tripathy[2], Dr. E Ravi Kumar[3], Aishwarya Pottabathula[4] , Yasaschandra Reddy Bandi[5] , Maadhurya Sri Surepalli[6] , Pujitha Dhoragolla[7]
[1,2,4,5,6,7] VNR Vignana Jyothi Institute of Engineering & Technology, Hyderabad, Telangana, India
[3] Vardhaman College of Engineering, Hyderabad, Telengana
[1]manmathnath.das@gmail.com, [2]jyotsnarani_t@vnrvjiet.in, [3]ravikumar.e@gmail.com,
[4]aishwaryapottabathula9@gmail.com, [5]yasaschandrareddy.bandi2003@gmail.com,
[6]maadhuryasrisurepalli99@gmail.com, [7]dhoragollapujitha123@gmail.com

Abstract
The proposed project looks forward to introducing a unique system for computer mouse control through hand gestures and voice commands with the use of advanced computer vision and speech recognition technologies. Traditional input devices are often limited in flexibility and accessibility, especially to people with physical disabilities and vision impairments. Our project will overcome these restrictions by allowing intuitive gesture-based and voice-based interaction. Our system uses advanced image processing and machine learning algorithms in real time to identify hand gestures so that users can command their computer using a series of hand movements sensed by a webcam or other cameras. Voice recognition also gives the user commands like "turn up the volume" or "click" for complete hands-free computing.
***Keywords**: Gesture Recognition, Voice Recognition, Computer Vision, Accessibility, Real-time Interaction, Hands-free Computing*

INTRODUCTION

In an age of high technological mergers, traditional input devices such as keyboards and mouse pose great difficulties and conflicts for individuals suffering from physical disabilities, as well as impairments to their vision. Based on previous discussions, we are designing a system that will redefine how a person interacts with their computer by combining hand gesture recognition and voice command capabilities. This not only provides better accessibility but also introduces a frictionless and intuitive interface for hands-free computing. A system that includes high-tech computer vision to capture and decode real-time hand gestures from a webcam or other gadgets will enable voice recognition so that the user can complement the hand gestures by speaking the command to be executed to achieve a more efficient interaction. With its feature of inclusiveness, the project has the potential to empower diverse users while offering more independence and convenience in computing tasks.

Existing Work

Many studies have been directed toward improving hand gesture recognition systems to enhance human-computer interaction and accessibility. One of the most important studies proposed a hand gesture recognition system using convex hull and contour-based algorithms to detect finger and palm positions, achieving real-time accuracy with minimal computational resources [1]. Another approach use fingertip tracking for gesture-based mouse control, which shows the potential of vision-based input devices to replace traditional ones and allows for a more intuitive user interface [2]. Other efforts involved optimal real-time gesture recognition systems and histogram-based background removal techniques, which significantly improved recognition accuracy under varied lighting conditions, in order to overcome one of the biggest challenges of gesture-based interaction [3] In addition, real-time finger tracking with an interactive quiz interface allows gestures to be mapped to certain actions, which makes highly useful for educational purposes and interactive learning [4]. Another study highlighted the use of CVZone libraries that made hand tracking and gesture mapping easier, while also making the system more reliable and efficient for gesture recognition across various cases [5]. A hybrid model using CNNs and transformer encoders has also been found to improve gesture recognition drastically by capturing spatial-temporal features efficiently and mainly through attention mechanisms to grasp the complexity of tasks for greater accuracy [6]. From an accessibility perspective, one of the projects was the introduction of a gesture-controlled virtual mouse system for physically disabled people. The system integrates real-time gesture recognition with voice command mechanisms, thus increasing its applicability and inclusivity, making technology more accessible to people with mobility limitations [7]. Furthermore, the use of GANs for dataset augmentation has been important in improving the performance of the model by providing diverse samples for training thereby increasing the robustness of the system [8]. Mobile applications have been developed to enable gesture-based control with voice commands for augmentation.

This allows for more effective multitasking, extending the scope of gesture-based interfaces in real-world applications [9]. A low-cost solution for gesture recognition without expensive hardware, such as special gloves or sensors, was developed. The system was able to deliver accurate results while maintaining low costs and ensuring scalability using video datasets and transfer learning [10]. Studies have also been conducted by researchers on multimodal approaches that combine gesture recognition with other technologies, such as voice commands and touch-based interaction. This results in a much more fluid and natural user experience because users can easily change between modes of interaction based on needs and context [11]. In one of these projects, hand gestures were incorporated into smart home automation. This was used to enable users to control their devices intuitively. This reflects the capacity of gesture recognition systems to play important roles within the IoT ecosystem [12]. According to one research paper, "Python, using PyAudio, is an interpreted language that still allows real-time audio processing." It indicates performance optimizations through NumPy and SciPy, which reduce latency and make Python a suitable choice for such applications [13]. Another study presented a more advanced model that integrated deep learning techniques and spatial as well as temporal features for recognizing gestures, thus enabling a more accurate dynamic interpretation of gestures [14]. Finally, some studies have investigated merging gesture recognition and virtual reality for the first time, providing individuals with the freedom to interact with a virtual environment via natural gestures. This merging promises new opportunities for immersive applications derived from entertainment, education, or training [15]. In summary, these works together show how gesture recognition systems may establish a new avenue in human-computer interaction and highlight current trends of ensuring that such systems work accurately, are usable, and accessible so that they can be used appropriately in applications ranging from everyday computing to specialized tasks, such as helping individuals with disabilities or immersive virtual experiences.

Background Work

Hand gesture and voice command-based virtual mouse systems have been an increasingly researched area in recent years owing to the high priority that the promising Ness of an intuitive and accessible alternative to traditional input devices such as mice and keyboards has increased over the years.

The initial attempts at virtual mouse systems involved basic image processing techniques, such as contour detection, optical flow, and feature matching, in the tracking of hand movements and gesture interpretation. Even though such techniques provide an elementary approach to the subject, they suffer from low accuracy and responsiveness owing to dynamic lighting or complex backgrounds. Moreover, integration with voice commands is lacking in these systems, which means that their flexibility and accessibility are reduced. The revolution that deep learning has brought to the field is mainly in the form of Convolutional Neural Networks (CNNs) for gesture recognition. CNN's automatic feature extraction of the CNN enabled it to potentially identify specific hand shapes and movements from video frames, which led to the mapping of gesture-specific operations to mouse functions such as clicking, scrolling, and dragging. For example, media pipes and CVZone have made hand tracking and gesture mapping easier, and gesture-controlled applications have been built with virtual mice. Although CNN-based systems are effective in spatial recognition, the temporal aspect of hand gestures, especially continuous and complex movements, is challenging to represent. To overcome this limitation, hybrid models based on the fusion of CNNs with temporal sequence models, including RNNs, LSTMs, and Transformers, have been developed. They captured spatial and temporal features to allow real-time recognition of dynamic gestures and transitions among actions. This was mainly realized through the power of transformers based on self-attention mechanisms to allow the processing of entire sequences of gestures, thereby enabling a superior accuracy and scalability. Further integration with voice commands makes virtual mouse systems more functional and accessible. Systems can interpret natural language processing models as open applications, change the volume, or click on a certain item by merely speaking. With tools such as Google Speech-to-Text and Microsoft Azure Speech Services, voice recognition technologies have enabled a fully multi-modal interaction for users with physical disabilities or for hands-free control. Modern hand gesture-and voice command-based virtual mouse systems combine real-time tracking of the hand with speech recognition to provide an intuitive user experience. In such systems, tracking is possible owing to depth-sensing cameras, RGB cameras, or even webcams, which observe movements and gestures of the hands. Advanced algorithms are used to preprocess and segment the input data to ensure robust performance under varying conditions. Voice commands are integrated into the system using NLP frameworks, thereby allowing users to issue contextual instructions that complement gesture-based controls.

Although these systems have shown some significant promise, the performance achieved by these diverse user demographics in diverse environments under various hardware configurations remains inconsistent. Areas of current active research on these systems continue to include dealing with gesture ambiguity, speech accents, and delays in computation time. Other necessary future developments that would be relevant for the mainstreaming of such systems include guaranteeing scalability, affordability, and handling privacy of voice and video data. This promising direction for virtual mouse systems to become accessible and user-friendly is hand-gesture recognition in conjunction with voice-command integration. Innovations in these areas will redefine human-computer interaction, enabling users with disabilities, multitasking capabilities, and a more natural and immersive computing experience.

Proposed System

The proposed system offers an innovative approach to human-computer interaction through control by hand gestures and voice commands. This will offer a competent alternative or supplement to conventional input devices, such as the mouse and keyboard, thus allowing for an intuitive mechanism of interaction that is free of hands. Advanced computer vision and speech recognition technologies have been leveraged to provide an answer with the intention of offering seamless and uninterrupted real-time interactions.

For hand-gesture-based mouse control, the system starts by capturing real-time video feed via a standard webcam. The captured video is then processed using MediaPipe to detect the user's hand and identify 21 key landmarks. Depending on the detected landmarks, the system uses binary encoding and ratio-based calculations to identify gestures in real-time. The recognized gestures are then mapped to computer-specific actions such as cursor movement, clicks, and scrolling. To further increase accuracy and stability, the system also incorporates Kalman filtering to reduce noise and diminish the impact of fast hand movements. This would provide very smooth and reliable cursor control, even under challenging environments or dynamic conditions. In addition to gesture recognition, the system uses speech recognition to decode and execute the voice commands. Libraries such as Speech Recognition and pyttsx3 allow the user to give orders such as "move mouse" or "click" giving the flexibility and accessibility of interaction through the system. Thus, it opens an opportunity to extend the scope of a system's intuitive natural interface that a computer may easily allow users. This is also achieved through interaction feedback using audio and visual responses, which include giving audio prompts or on-screen notifications to confirm commands and gestures, thereby improving user experience. It is a form of feedback in interactions that is useful in making interactions more engaging and intuitive, and enhancing user confidence when using it. There is adequate testing in varying environments, light, and people using the system to establish reliability. Such an adaptation would ensure that the system operates in real-life circumstances. Plans involve adding multilingual support in voice commands, as well as with other human-computer interaction systems. This will fill the gap between technology and people, because it offers a natural, accessible, and fficient interface. Its deployment will empower people, particularly those with physical disabilities, by offering modern alternative input devices. The fact that the system combines gesture recognition with voice commands is another significant step forward toward an all-inclusive, intuitive computing process for all people.

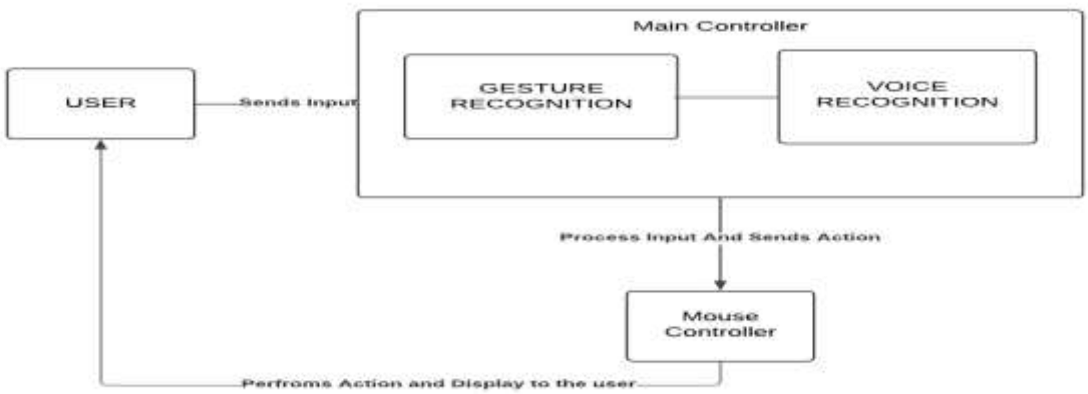

Figure 1 System Architecture

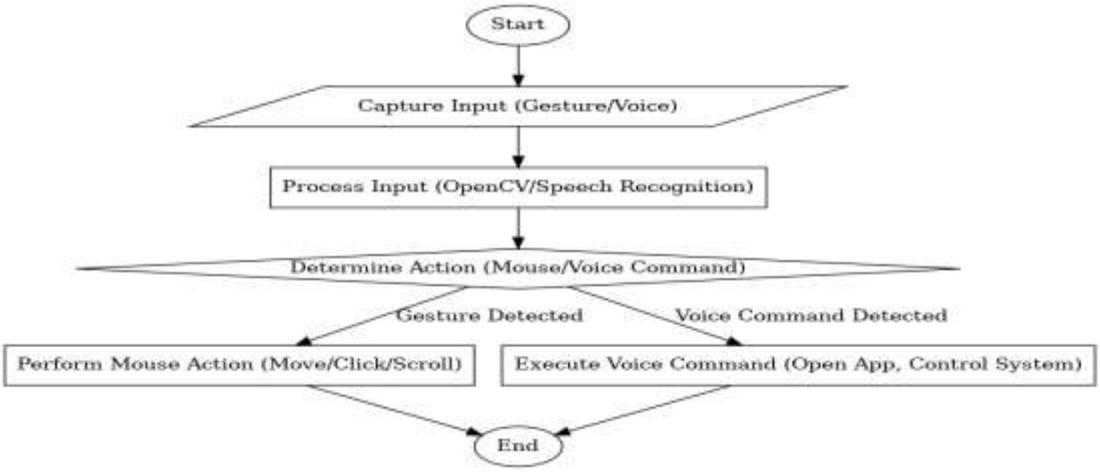

Figure 2 Flow of the proposed system

DESIGN AND METHODOLOGY

Planning and Requirement Analysis

The project begins with a thorough analysis of the needs of users, particularly in designing an accessible interface for people with physical disabilities or visual impairments. The key requirements of the system are to capture hand gestures using a webcam and interpret voice commands correctly. Feasibility studies were then conducted to assess the available hardware, including cameras and microphones; suitable software libraries and frameworks, including MediaPipe, OpenCV, and speech recognition. Finally, a roadmap is defined for stages, such as development, testing, and deployment.

System Design

Architecture Overview: The system is designed such that computer vision and voice recognition can be perfectly integrated, as follows:

Input Layer: Video feed is captured by a webcam for gesture recognition, and the voice commands are recorded by a microphone.

Processing Layer: MediaPipe extracts hand landmarks, and complex algorithms, such as binary encoding and ratio-based calculations, are used for gesture recognition. Voice commands were executed using speech recognition libraries.

Output Layer: Recognized gestures and voice commands are assigned to mouse movements; feedback occurs with visual feedback, such as on-screen display and audio with text-to-speech confirmation.

Implementation

For gesture recognition, the MediaPipe algorithm was applied to the retrieval of real-time hand landmarks from video feed. Then, advanced algorithms identify the distance, angles, and ratio between the landmarks for specific gestures. This can be done with the help of Equations (1) and (2). Finally, the gestures are mapped to mouse actions, such as the movement of the cursor, clicking, and scrolling.

For voice recognition, speech commands are captured using libraries such as speech recognition. Then the recognized commands such as "scroll down" or "click" are mapped to predefined actions.

Integration of both systems: Both modules-gesture and voice–are integrated within the system. Concurrency leads to a technique called threading, whereby both can simultaneously come and give no delays for a period.The feedback Mechanism will be performed for both systems as follows: For a hand gesture recognition mouse, the GUI will have the text displayed on it with respect to the action performed. For voice recognition, visual feedback will also result through an auditory device using a tool such as text-to-speech on successful completion.

esting and Evaluation

Lighting Conditions: Testing under different lighting conditions that prove stability and precise recognition in this context. User Variations: Tests were conducted with different users to check for the adaptability of the system towards different hand shapes and speech patterns. Error Handling: False gestures or commands caused by environmental noise are detected and ignored through mechanisms.

System Performance Evaluation: By recognition of accuracy and speed and response time. and also through robustness under diverse conditions.

Testing focuses on the accuracy, speed, and robustness of the system to ensure that it meets user expectations.

Deployment and User training

The system was tested and validated once and then packaged in a user-friendly interface for deployment. Next, documentation of the installation and setup is created. An optional training guide helps users to adapt to the new interaction paradigm, making hands-free computing as smooth as possible.

Output and Evaluation

The system allows hands-free control of mouse actions and computer functions by using gestures and voice commands. Outputs include mouse cursor movements, clicks, scrolling, and opening applications. Evaluation can be measured through recognition accuracy and response time, user satisfaction, and other aspects of user feedback, particularly for users with disabilities or vision deficiencies, which are collected to assist in improving the system.

This systematic methodology ensures that the project is efficiently developed, thereby achieving its goals of accessibility and innovation.

Equations

$$\text{Distance} = \sqrt{(x2 - x1)^2 + (y2 - y1)^2} \qquad (1)$$

$$\text{Angle} = \cos^{-1}\left(\frac{AB^2 + BC^2 - AC^2}{2 \cdot AB \cdot BC}\right) \qquad (2)$$

Results

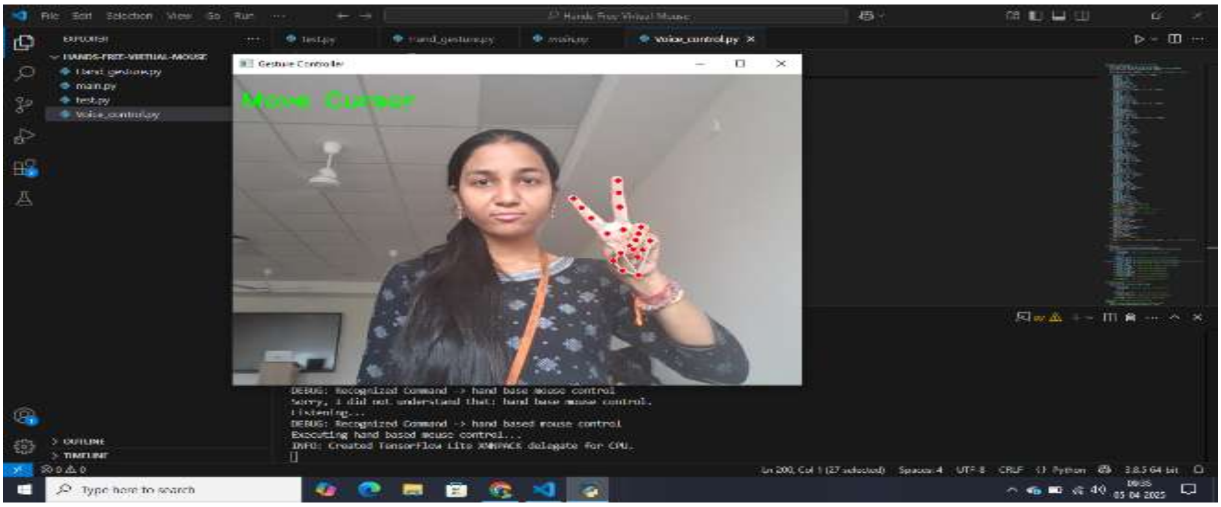

Figure 3 Moving the cursor using index and middle finger by making v shape.

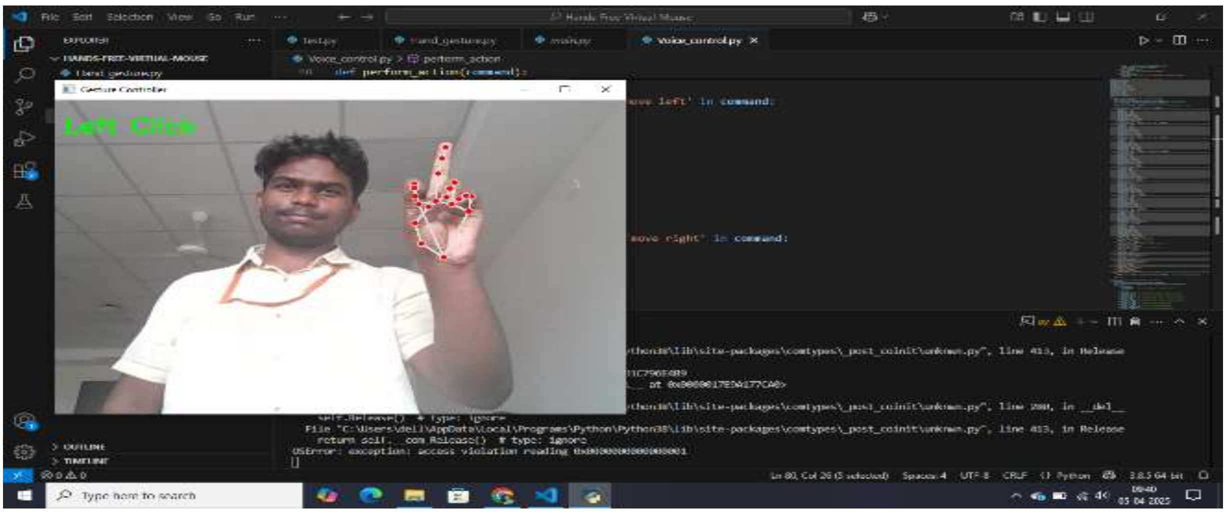

Figure 4 Left clicking the mouse by bending index finger..

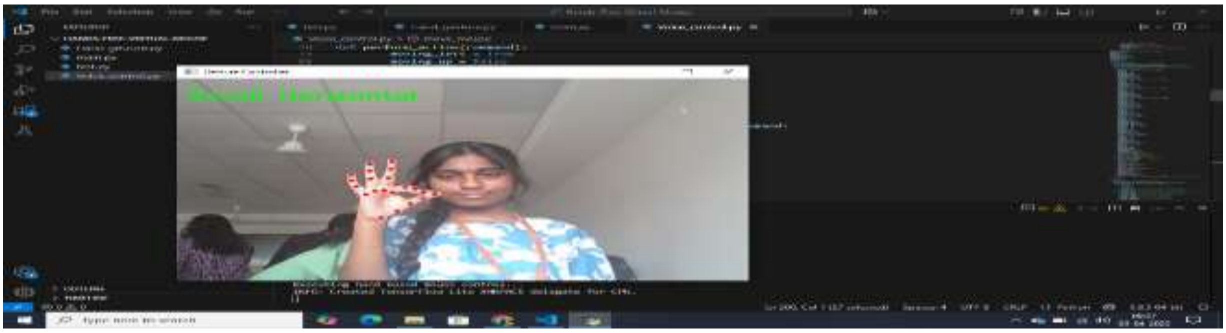

Figure 5 Scrolling horizontally by pinching right hand's index and thumb finger and moving them horizontally

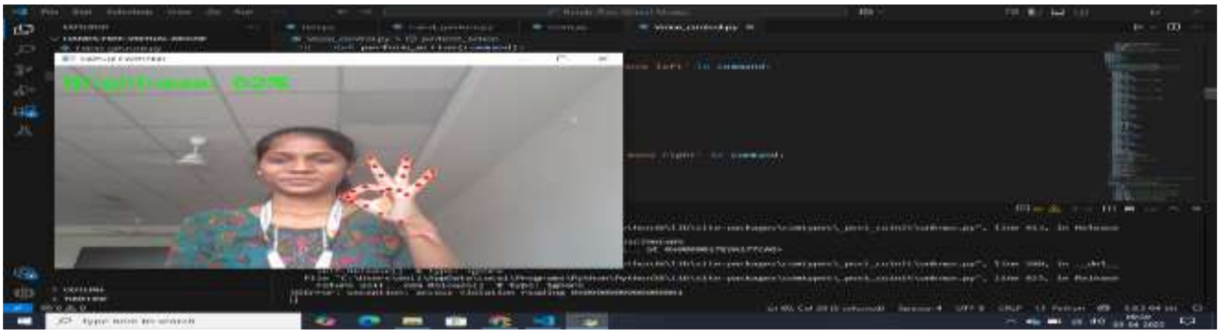

Figure 6 Changing the brightness by pinching left hand's index and thumb finger and moving them horizontally

The system allows easy control of the computer mouse with simple hand movements. Figure 3 shows movement of the cursor, which is done by making a "V" shape using the index and middle fingers so that one can navigate precisely. Figure 4 shows a left-click operation, done by curving the index finger while holding the rest of the hand still. Figure 5 illustrates horizontal scrolling, which is activated by pinching the index and thumb fingers of the right hand together and sliding them horizontally. Figure 6 illustrates brightness control, where the same pinching action with the index and thumb fingers of the left hand and subsequent horizontal sliding adjusts the brightness of the screen accordingly. These gestures are an effortless and accessible substitute for conventional mouse input.

Figure 7 Enabling voice-controlled mouse using voice command

Figure 8 Moving mouse using voice command


Figure 9 Scrolling using voice command

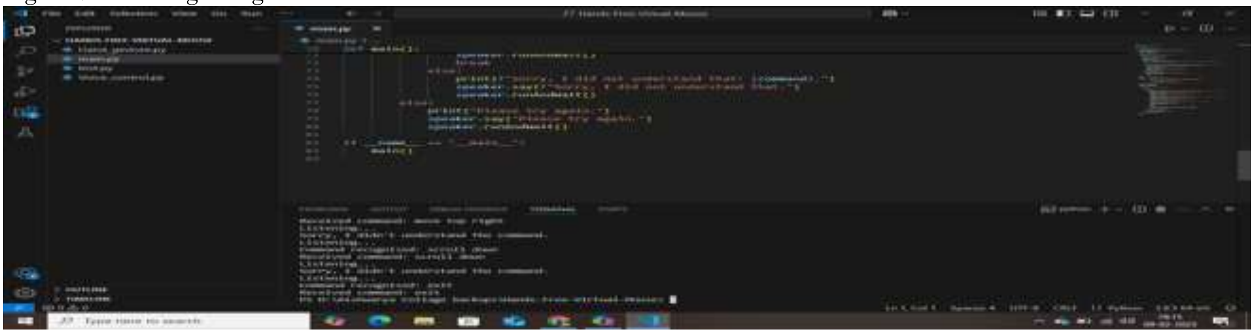
Figure 10 Exiting voice-controlled mouse using voice command

The system also provides voice commands to control the mouse hands-free, improving accessibility and convenience. Figure 7 displays the voice control of the mouse by a custom voice command for activating the voice interaction mode. After activation, users can interact with the computer using simple vocal commands. Figure 8 shows moving the mouse based on directional voice commands like "move up" or "move left." Figure 9 displays scrolling behaviour, with users being able to utter commands like "scroll down" or "scroll right" for scrolling through content. Last but not least, Figure 10 indicates how the system can be terminated by giving a clear voice command like "exit," so that users can close the session gracefully. This voice-control provides an accessible solution for those who might have limited mobility or simply prefer to use verbal interaction.

Table 1 Comparison Chart of Existing models

S. No	**Parameters**	**Gesture controlled virtual mouse and voice automation with integrated gesture database[17]**	**Gesture Controlled Virtual Mouse with Voice Assistant[16]**
1	Models Used	CNN, DNN, Hand gesture recognition systems etc.	Python modules
2	Accuracy	80-93.25%	75-85%
3	Reliability	Reliable in stable lighting conditions	Less reliable in noisy environments
4	Performance	Fast and highly flexible	Moderate and less audio recognition in noisy environments
5	Data set used	Large dataset of hand gesture	Python libraries

Table 2 Ratings of our proposed model based on user experiences

S. No	**Parameters**	**Gesture-only**	**Voice-only**	**Overall Integrated system**
1	Accuracy	94%	90%	92%
2	Accessibility	Easily accessible using intuitive and familiar hand gestures.- 89%	Easily accessible using simple and commonly used voice commands – 90%	Combines both input modes, offering flexible and inclusive interaction. – 90%

3	Reliability	Performs well in good lighting and with steady hand movement, works well with single person. – 88%	Works reliably in quiet environments; may struggle with noise or unclear speech. – 93%	Highly reliable, ensuring consistent performance across various environments – 91%
4	Responsiveness	90%	85%	88%
5	Interface	95%	90%	93%

CONCLUSION

The Hands-Free Mouse with Gesture and Voice Control projects offer an innovative shift in the way people communicate with technology. Through the use of gesture recognition and voice control, it offers an intuitive and hands-free form of exploring the virtual world, thus eliminating the need to use traditional input devices, such as a mouse and keyboard. This system not only promotes increased user ease, but also serves as a milestone of significant importance in terms of accessibility. Individuals with physical disabilities, particularly those with low mobility, are now able to interact more seamlessly with technology, which allows them to access computers and digital interfaces without any difficulty. As the system develops further, the possibilities for its expansion become immense. Adding support for multilingual voice commands would enable the system to serve a more globalized population of users. Moreover, increased developments in machine learning and computer vision would make gesture recognition more accurate and reliable, even under less optimal conditions, such as low light or excessive motion. In summary, the Hands-Free Mouse with Gesture and Voice Control project represents the future of human-computer interaction, where technology adjusts to the user and not vice versa. Through the use of sophisticated technologies such as computer vision and speech recognition, the system creates new avenues for more inclusive and user-friendly interactions. Their potential to improve usability and access is particularly strong for individuals with physical disabilities or impairments. As the system evolves and more capability is added, it may become an essential utility across a broad spectrum of applications ranging from healthcare and education to entertainment and everyday online life. This project not only contributes to the evolution of user interfaces but also paves the way for a technologically accessible and adaptive world that encourages empowerment and autonomy for all.

Future Scope

Hands-free mice with gesture and voice control projects have immense potential for future advancement, particularly in enhancing the accuracy and adaptability of gesture and voice recognition. By integrating more advanced deep learning models, the system can improve its ability to recognize a broader range of hand gestures, thereby enabling more precise and efficient control. Future iterations could also incorporate more complex gestures such as multi-finger tracking and customizable commands, allowing users to personalize their interactions based on their needs. In addition, improving the speech recognition module by supporting multiple languages, accents, and personalized voice commands would make the system more inclusive, enabling global accessibility and ease of use. Another important area of development is enhancing the adaptability of the system to different environments and conditions. Currently, gesture recognition may face challenges in low-light settings or with high background noise; however, future improvements could incorporate advanced computer vision techniques and noise-cancelling algorithms to overcome these limitations. The system can also be optimized to function effectively across various devices such as smartphones, tablets, and smart TVs, broadening its applications beyond traditional computing. Moreover, integrating AI-driven predictive modelling can help the system learn user preferences over time, offering a more intuitive and personalized experience. Beyond its individual use, this technology has significant applications in various industries, including healthcare, gaming, and smart home automation. In healthcare, the system can be further refined to assist individuals with severe mobility impairments in the effortless control of assistive devices. The gaming industry could leverage the project's capabilities to create more immersive experiences through gesture-based controls, whereas smart home integration would allow users to interact with appliances and digital assistants hands-free. As advancements in AI, machine learning, and sensor technology continue, this project has the potential to revolutionize human-computer interaction, paving the way for a more accessible and futuristic digital world.

Conflicts of Interest

The authors have no conflicts of interest to declare.

REFERENCES

[1] ."Hand Gesture Recognition Using OpenCv and Python" by V. Harini, V. Prahelika, I. Sneka & P. Adlene Ebenezer in 2020.
DOI : https://doi.org/10.1007/978-3-030-41862-5_174
[2] ."Vision-Based Interpretation of Hand Gestures for Remote Control of a Computer Mouse" by Antonis A. Argyros and Manolis I.A. Lourakis
DOI : https://doi.org/10.1007/11754336_5
[3] ."Mouse Cursor's Movements using Voice Controlled Mouse Pointer" by Saritha and Kiran Kumar Kaki in 2013.
DOI : doi=45fd38ade1fc6fa5d3f5cce4018175cce949059e
[4] ."Real-time Audio Recognition for Hearing Impaired" by Muralidhar Peddi; S H V Harsha Vardhan; Giridhar Peddi; Satwik Ponnam; C Rani; M Rajesh Kumar in 2024.
DOI : 10.1109/AIIoT58432.2024.10574757
[5] ."Hand Gesture Recognition Based on Computer Vision: A Review of Techniques" by Munir Oudah , Ali Al-Naji and Javaan Chahl in 2020.
DOI : https://doi.org/10.3390/jimaging6080073
[6] ."Gesture-Based Virtual Mouse Interface" by Thakkuri Sanjay, Nagampeta Ashish Reddy, Puskuri Chinku, Kancha Vamshi Vardan, Yatin Shukla in 2024.
DOI : https://doi.org/10.56726/IRJMETS50946
[7] ."Mouse Cursor Control System Based on Hand Gesture" by Horatiu-Stefan Grif , Cornel Cristian Farcas in 2016.
DOI : https://doi.org/10.1016/j.protcy.2016.01.137
[8] ."Human Hand Gesture Based System for Mouse Cursor Control" by Horatiu-Stefan Grif and Traian Turc in 2018.
DOI : 10.1016/j.promfg.2018.03.147
[9] ."Real-time virtual mouse system using RGB-D images and fingertip detection" by Dinh-Son Tran, Ngoc-Huynh Ho, Hyung-Jeong Yang, Soo-Hyung Kim & Guee Sang Lee in 2020.
DOI : https://doi.org/10.1007/s11042-020-10156-5
[10] . "Research Progress of Human–Computer Interaction Technology Based on Gesture Recognition" by Hongyu Zhou , Dongying Wang ,Yang Yu and Zhenrong Zhang in 2023.
DOI : https://doi.org/10.3390/electronics12132805
[11] . "Control Mouse and Computer System Using Voice Commands" by Mahammad Rafi , Khan Sohail Ahmed , Shaikh Huda , Lutful Islam in 2016.
DOI : https://ijret.org/volumes/2016v05/i03/IJRET20160503077.pdf
[12] . "A Danish Voice-Controlled Internet Browser for Motor-Handicapped Users" by Tom Brøndsted & 2 Erik Aaskoven.
DOI : https://www.isca-archive.org/interspeech_2005/brndsted05_interspeech.pdf
[13] . "Development of Real-Time Audio Applications Using Python" by Yuri De Pra, Federico Fontana and Michele Simonato.
DOI : https://www.researchgate.net/profile/Federico-Fontana-6/publication/331899778_DEVELOPMENT_OF_REAL-TIME_AUDIO_APPLICATIONS_USING_PYTHON/links/5e68baef4585153fb3d60090/DEVELOPMENT-OF-REAL-TIME-AUDIO-APPLICATIONS-USING-PYTHON.pdf
[14] . "Gesture controlled human–computer interface for the disabled" by Oskar Szczepaniak Dariusz Sawicki in 2017.
DOI : 10.13075/mp.5893.00529
[15] . "A Real-Time Human-Computer Interaction" by Pei Xu in 2017.
DOI : https://doi.org/10.48550/arXiv.1704.07296
[16] . "Gesture controlled virtual mouse with voice assistant" by Ismail Khan, Vidhyut Kanchan in 2024.
DOI : 10.47857/irjms.2024.v05i01.0127
[17] . "Gesture controlled virtual mouse and voice automation with integrated gesture database" by H. S. Annapurna, Koushik Umesh Pai, Likhith Gowda M J, Jitendra Patel N B, Kushala H Ein 2024.
DOI : https://ijcrt.org/papers/IJCRT24A5813.pdf