Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 15, Issue 3, April 2024: 12-19
DOI: 10.53555/tojqi.v15i3.10282

Research Article

Gesture Voice: Revolutionizing Human-Computer Interaction with an AI-Driven Virtual Mouse System

Dr. Nisha Patil[1]*, Moeeza W. Ansari[2], Sakshi R. Jadhav[3], Manjeeri G. Mhaske[4], Rohit K. Kaurani[5]

Abstract

In this paper, we propose an innovative AI-driven virtual mouse system that seamlessly integrates hand gestures and voice commands for intuitive interaction with computing devices. Traditional input methods such as physical mice and keyboards can be limiting for users with mobility impairments or in scenarios where hands-free operation is desired. Our system addresses these challenges by leveraging advancements in computer vision and natural language processing to enable users to control the mouse cursor and execute commands through hand gestures and voice input.

The core of our system lies in the fusion of hand gesture recognition and voice assistant technologies, providing users with a versatile and efficient means of interacting with computers and other smart devices. Through a combination of depth sensing cameras and machine learning algorithms, the system accurately detects and interprets a wide range of hand gestures, allowing users to perform actions such as cursor movement, clicking, dragging, and scrolling with ease.

Furthermore, security is a paramount concern in modern computing environments, especially with the increasing prevalence of AI-driven interfaces. To address this, our system incorporates robust security measures, including user authentication through voice biometrics and encryption of communication channels to safeguard sensitive data.

Overall, our AI virtual mouse system represents a significant advancement in human-computer interaction, offering a seamless and secure interface that empowers users to interact with computing devices more naturally and efficiently.

***Key Words*:** AI virtual Mouse System, Human-Computer Interaction, Computer Vision, Gesture Recognition, User Authentication

1. INTRODUCTION

Artificial Intelligence refers to the simulation of human intelligence processes by machines. These processes include learning, problem-solving and Expert Systems.AI can enable the creation of virtual mouse that can be controlled through various means, including gestures or voice commands. The virtual mouse is particularly useful for individuals with disabilities who may have difficulty using traditional physical mouse which is used by Adaptive Interface. Histograms of Oriented Gradient (HOG) are descriptors rotationally invariant which have been used in optimization problems as well as in computer vision [1].

In our case, we apply in the face detection problem in the traditional method, users interact with the system using a physical input device such as a mouse and trackpad, AI enabled virtual mouse systems can leverage diverse input modalities such as gesture recognition and voice control. The traditional method relies heavily on the hardware components of the input device. AI-powered virtual mouse systems have the potential for automation and intelligent assistance features. Computer vision

[1]*Assistant Professor, Guru Gobind Singh College of Engineering and Research Centre, Nashik, Maharashtra, India
[2]Student, Guru Gobind Singh College of Engineering and Research Centre, Nashik, Maharashtra, India
[3] Student, Guru Gobind Singh College of Engineering and Research Centre, Nashik, Maharashtra, India
[4] Student, Guru Gobind Singh College of Engineering and Research Centre, Nashik, Maharashtra, India
[5] Student, Guru Gobind Singh College of Engineering and Research Centre, Nashik, Maharashtra, India

algorithms can identify and track objects in real-time. hand gestures, body movements, or to enable gesture-based interaction with the system. The human hand is capable of transmitting large amounts of information via typing or sign language. It can form a great number of complexes poses and is different from person to person in terms of size, and shape. [2]

Gesture recognition is a technology enabling computers to interpret human gestures for device control or system interaction, relying on data acquisition and pattern recognition. Applications include sign language recognition, security, and smart TVs. Desktop-based voice assistants, which recognize human voices and respond through integrated systems, are also prevalent. APIs are utilized to convert text to audio. Python programming and artificial intelligence are employed in our project, utilizing a microphone for input and a speaker for output.

2. PROBLEM STATEMENT

The project aims to develop an AI-driven virtual mouse system that relies on hand gestures and integrates a voice assistant for enhanced user interaction. The primary challenge is to create a robust and reliable hand gesture recognition system that accurately translates a user's hand movements into cursor control, enabling them to navigate and interact with a computer without physical peripherals.

Additionally, incorporating a voice assistant adds a layer of complexity, as it requires the seamless integration of natural language processing and speech recognition technologies to enable users to execute commands and interact with applications through voice commands. Ensuring security is another critical aspect, as unauthorized access or control of the system poses a significant risk. The project will address security concerns by implementing user authentication and access control mechanisms, which is essential to protect sensitive data and maintain the privacy and integrity of the system.

The success of our work will enhance the accessibility and usability of computers for a broad range of users, including those having physical disabilities, while simultaneously addressing potential security vulnerabilities associated with such systems.

3. OBJECTIVES

The newly developed AI virtual mouse for gesture recognition which allows computers to interpret human gestures as input for interacting with systems. Following are the objectives of the research topic:

- Introducing an innovative AI-driven virtual mouse system that integrates hand gestures and voice commands for intuitive interaction with computing devices, aiming to overcome limitations associated with traditional input methods such as physical mice and keyboards, especially for users with mobility impairments or in hands-free operation scenarios.
- Leveraging advancements in computer vision and natural language processing to enable users to control the mouse cursor and execute commands through hand gestures and voice input.
- Fusion of hand gesture recognition and voice assistant technologies at the core of the system to provide users with a versatile and efficient means of interacting with computers and smart devices, including actions such as cursor movement, clicking, dragging, and scrolling, facilitated by depth sensing cameras and machine learning algorithms.
- Addressing security concerns in modern computing environments, particularly with AI-driven interfaces, by incorporating robust security measures such as user authentication through voice biometrics and encryption of communication channels to safeguard sensitive data.
- Demonstrating a significant advancement in human-computer interaction through the AI virtual mouse system, offering a seamless and secure interface that empowers users to interact with computing devices more naturally and efficiently.

4. LITERATURE SURVEY

Computer vision algorithms can identify and track objects in real-time. hand gestures, body movements, or to enable gesture-based interaction with the system.
Gesture recognition is a technology that allows computers to interpret human gestures as inputs for controlling devices or interacting with systems. Which Work on Data Acquisition and Pattern Recognition Some applications are Sign Language Recognition, Security, Smart TVs.
The methods of communication utilized to operate computers notably impact human-computer interactions. Gesture control or eye movement tracking systems can potentially substitute or trigger certain functions typically performed by a mouse or keyboard [3]. A gesture-based interface enables users to interact with a computer via fingertip detection in RGB-D inputs. Initially, the hand region and palm center are identified from depth images provided by the Kinect V2 skeletal tracker, then converted to binary images. Subsequently, hand contours are extracted and analyzed using a border-tracing algorithm [4].
MediaPipe Hands is a high-quality finger and hand tracking system that employs machine learning (ML) to detect 2D and 3D landmarks of a hand from a single image. achieves real-time performance on mobile phones for multiple users [5]. Table 1. shows the advancements in the era of AI Virtual mouse system.

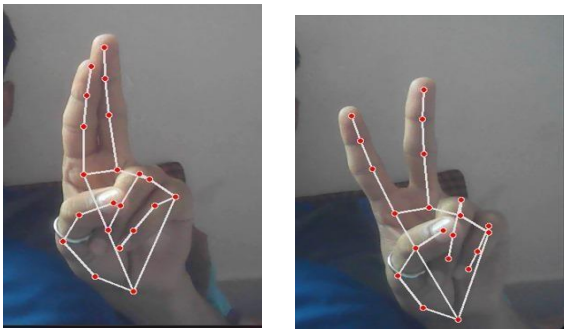
Table 1. Advancements in the era of AI Virtual mouse system

SR. NO.	DATE	REFERENCES	THESIS	ADVANCEMENT
1	2023	[6]	AI Virtual Mouse System Using Hand Gesture and Voice Assistant	High-quality finger and hand tracking system, Machine learning (ML) is used to determine 2D and 3D landmarks of a hand from a sing
2	2023	[7]	High-quality finger and hand tracking system, MediaPipe Hands. Machine learning (ML) is used to determine 2D and 3D .	The OpenCV library is used to analyze data from photos and videos, including face and object detection.
3	2020	[8]	The OpenCV library is used to analyze data from photos and videos, including face and object detection	To perform Object detection in Aerial Images, Convolutional Neural Networks,
4	2023	[9]	Hand Gesture And Voice Assistant	The modules NumPy and Python will be utilised to construct this system
5	2023	[10]	AI voice assistant for smartphones with NLP technique	AI voice assistant for smartphones with NLP technique
6	2022	[11]	Hand gesture and voice- controlled mouse for physically challenged using computer vision	A Convoluted Neural Network (CNN) efficient for gesture recognition.
7	2021	[12]	Security System by Face Recognition Using	Using Independent Component Analysis (ICA) and Principal Component Analysis (PCA).

5. METHODOLOGY

5.1 Hand Gesture Tracking

The Utilize computer vision techniques like Convolutional Neural Networks (CNNs) or hand-crafted feature extraction methods to detect and recognize hand gestures. Convolutional Neural Networks (CNNs) are a notable approach that has shown considerable potential in optimizing SCM systems [13]. Techniques such as background subtraction, skin color segmentation [14], or deep learning-based hand pose estimation can be used. This program is a gesture controller that uses the MediaPipe library for hand tracking and recognition. It allows the user to control various actions on the computer using hand gestures captured by the camera. Let's break down the main components of the program. Figure 1(a). shows to capture a finger tip.



(a) (b)

Figure 1. Capturing Fingertips using the webcam & Gesture for the computer to perform right button click

Gesture for the computer to perform scroll up function. successful mapping and extraction of distinct hand features. The suggested system is designed to perform a variety of operations,including volume control, right- and left- clicking, double- clicking, drag-and-dropping, multiple item selection, etc. Figure 1(b). performs a task on virtual system for right click.

Recognition and execution of intentional hand movements are facilitated by the Hand Gesture Recognition and Execution system. This system utilizes both the Mediapipe and OpenCV modules. Hand actions can be predicted based on the number and types of fingers present. For instance, a hand gesture with five fingers is categorized as neutral, while a gesture with two fingers indicates a cursor movement, and a gesture with the right finger pulled down signifies a right-click action shown in Figure 3 [15].

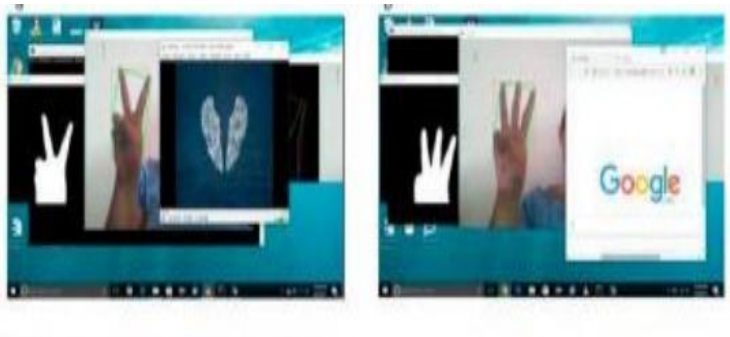


Figure 3. Cursor Control System [15]

5.2 Voice assistant Integration

Incorporate a voice assistant module using speech recognition algorithms to enable voice commands for additionalcontrol functionalities. Use libraries like Google SpeechRecognition or CMU Sphinx for this purpose. Design an intuitive user interface that displays visual feedback for hand gestures and voice commands. Provide clear instructions and feedback to the user to enhance usability. Figure 4. shows to recognize voice commands like as (Siri & Google Assistant)

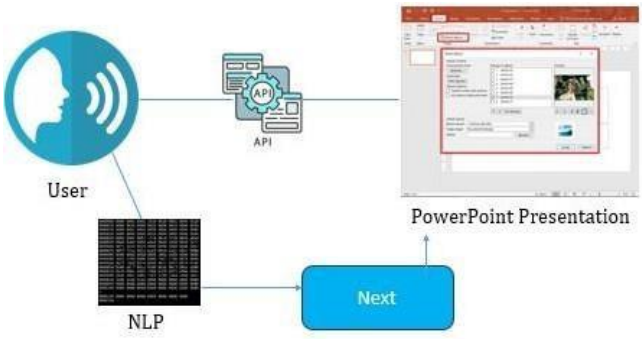


Figure 4. Voice Assistant

In the first critical stage, voice assistants can hear all instruction from user. Theyuse advanced audio processing techniques to meticulously analyze the user's voice commands captured through the microphone. Figure 5. shows natural language processing into machine language using (Voice-Based system).

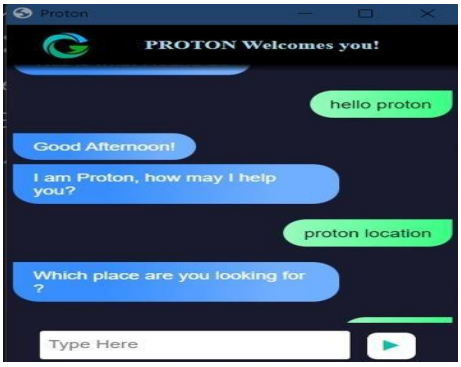


Figure 5. PROTON Voice-Recognition

This analysis goes beyond simply understanding the words. It considers factors like tone, emphasis, and even background noise. By dissecting these elements, the assistant can distinguish between a request and casual conversation, and even identify theuser's intent. This initial phase paves the way for the magic to happen - whether it's gathering information from the real world through a web search, controlling the device's internal data like music playlists, or even automating tasks through connected smart home devices. It's the crucial first step in the voice assistant's dance, translating spoken instructions into concrete actions. Figure 6. Shows the task with execute commands.

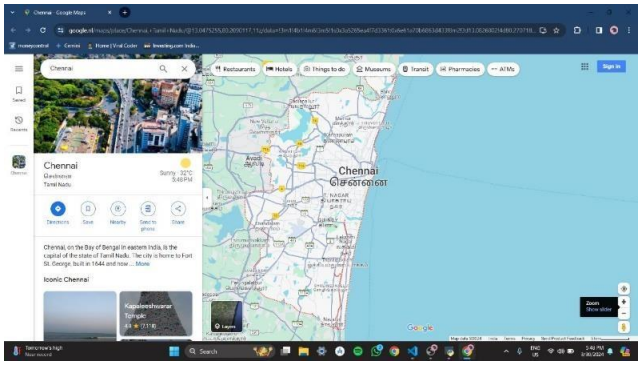


Figure 6. Catch a voice to perform a Task

To execute a task using a Voice-Based system, such as instructing a virtual assistant like Proton to search for information like "Where is Chennai," incorporating voice recognition and natural language processing (NLP) technologies. This enables users to control the virtual mouse system through voice commands, thereby enhancing accessibility and search result accuracy.

5.3 Security Features

Implement security measures such as facial recognition for userauthentication before granting access to the system. Utilize computer vision-based facial recognition algorithms like OpenCV's Haar cascades or deep learning-based approaches using frameworks like TensorFlow or PyTorch. Conduct thorough testing and evaluation of the system's performance, including accuracy, responsiveness, and security. Collect feedback from users to identify areas for improvement anditerate on the design accordingly. Figure 7. Shows face detection for authorized user.

To enhance prediction accuracy beyond that achievable by individual models, a collection of learning models is combined through a technique known as ensemble learning [16]. Single-factor authentication (SFA) relies solely on the user's possession of certain knowledge. While password-based authentication is suitable for accessing applications or websites, it is not highly secure for online

financial transactions [17]. Algorithmic processes enabled by AI-powered machine learning assist in real-time risk identification [18].

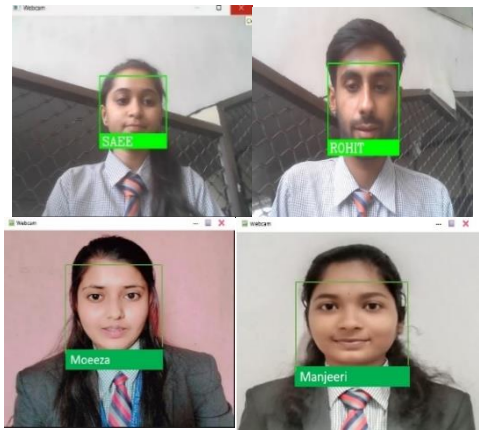


Figure 7. Face Detection

6 SYSTEM ARCHITECTURE

Our work involves creating an AI-driven virtual system capable of responding to hand gestures and voice commands, supported by diagrams and algorithms. Our objective is to transform human-computer interaction by introducing an innovative AI-powered Virtual Mouse System that integrates hand gesture recognition and voice assistance. Conventional input methods such as keyboards and computer mice can be restrictive, particularly for users with mobility impairments or in situations where touchless control is preferred. Our system empowers users to engage with their computers in a more natural and intuitive manner.

To execute tasks and analyse the system, we employ a combination of multiple techniques within our software. The virtual-based system is designed to recognize and execute commands. Visual representations illustrate the system's design and appearance. An example application could involve controlling a PowerPoint presentation through hand gestures and voice commands. The system boasts real-time responsiveness, enhancing user interaction and efficiency. Figure 8. shows architecture of system software.

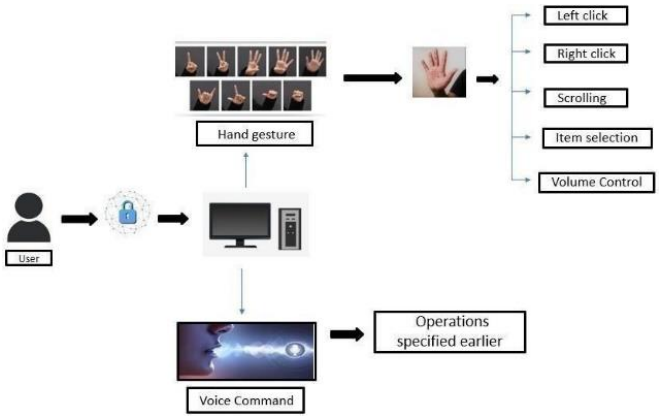


Figure 8. System Architecture and Design

CONCLUSION

The integration of AI-driven hand gesture and voice assistance technologies with robust security measures marks a significant advancement in human-computer interaction. This sophisticated system not only enhances user convenience but also ensures the protection of sensitive information. As technology continues to progress, this innovative fusion sets the stage for a future where human-machine communication is more intuitive, secure, and efficient than ever before.

By creating an AI-powered virtual system capable of responding to hand gestures and voice commands, we aim to revolutionize human-computer interaction, providing a more inclusive and natural experience for users. Through a combination of advanced techniques and real-time responsiveness, our system opens new possibilities for interaction and control, offering a glimpse into the future of computing.

REFERENCES:

[1] W. Schwartz, A. Kembhavi, D. Harwood, and L. Davis.Human Detection Using Partial Least Squares Analysis.2009

[2] Križnar V, Leskovšek M, Batagelj B. editors. Use of Computer Vision Based Hand Tracking in Educational Environments: 44th International Convention on Information, Communication and ElectronicTechnology (MIPRO), 27 September 2021 - 01 October2021, Opatija. Croatia: IEEE; 2021

[3] Katona, J. A Review of Human–Computer Interaction and Virtual Reality Research Fields in Cognitive InfoCommunications. *Appl. Sci.* 2021, *11*, 2646. https://doi.org/10.3390/app11062646.

[4] Hyung-Jeong Yang, et, al. Real-time virtual mouse system using RGB-D images and fingertip detection in Multimedia Tools and Applications, 2020, https://doi.org/10.1007/s11042-020-10156-5.

[5] J Kumara Swamy, et, al, AI Based Virtual Mouse with Hand Gesture and AI Voice Assistant Using Computer Vision and Neural Networks, Journal For Research in Applied Science and Engineering Technology, 2023.

[6] Likitha R, et. al. International Journal of Engineering Research and Applications www.ijera.com ISSN: 2248-9622, AI Virtual Mouse System Using HandGestures and Voice Assistant Vol. 12, Issue 12, December 2022, pp. 132-138

[7] Mr. E.Sankar, B.Nitish Bharadwaj, A.V.Vignesh Assistant Professor, Dept. Of CSE, SCSVMV Sankar CHAVALI Sri Chandrasekhar Endra Saraswathi Viswa Mahavidyalaya Universityr Virtual Mouse Using Hand Gesture this publication (2023) at: https://www.researchgate.net/publication/372165002

[8] Nisha Patil, Niketan Bothe, Shivam Gulve, “Image Scene Understanding -Object Detection in Aerial Images using Convolutional Neural Networks”, International Journal for Scientific Research & Development, Vol. 8, Issue 10, 2020

[9] Aarti Morajkar , Atheena Mariyam James , Minoli Bagwe , Aleena Sara James (2023)Hand gesture and voice-controlled mouse for physically challenged using computer vision

[10] Fungai Jacqueline Kiwa, Tavengwa Masamha , Shakemore Chinofunga - 979-8-3503-1812-8/23/$31.00 ©2023 IEEE - AI voice assistant for smartphones with NLP techniques Fungai Jacqueline Kiwa ICT and Electronics.

[11] Dian Christy Silpani, Keishi Suematsu (2022) A Feasibility Study on Hand Gesture Intention Interpretation Based on Gesture Detection and Speech Recognition

[12] Aisha Bazama , Fawzia Mansur* , Nura Alsharef Published: 09-06-2021 Security System by Face Recognition:http://creativecommons.org/licenses/by/4.0

[13] Patil, N., Bhise, A., & Tiwari, R. K. (2024). Fusion deep learning with pre-post harvest quality management of grapes within the realm of supply chain management. *The Scientific Temper*, *15*(01), 168–176. https://doi.org/10.58414/SCIENTIFICTEMPER.2024.15.1.23

[14] Patil, Nisha, Archana Bhise, and Rajesh Kumar Tiwari. "GRAPE QUALITY PREDICTION IN PRE-POST HARVESTING WITH IMPLEMENTATION OF FUSION DEEP LEARNING." *Journal of Analysis & Computitions* (2023).

[15] Sundus Munir, Shiza Mushtaq, Afrozah Nadeem, SyedaBinish Zahra Hand Gesture Recognition: A Review

[16] S. Gore, Y. Nagalakshmi, P. Knowles, K. G. Gupta, N. S. Jagtap and R. P. Sali, "Improvised Ensemble Model for Fast Prediction of DoS/DDoS Attacks in Various Networks," *2023 1st International Conference on Cognitive Computing and Engineering Education (ICCCEE)*, Pune,

India, 2023, pp. 1-5, doi: 10.1109/ICCCEE55951.2023.10424447.

[17] Kaustubh Chude, et., al," Multi-factor Authentication for Physical Access", International Journal of Scientific Research in Engineering and Management (IJSREM), Volume: 06 Issue: 05, May – 2022.

[18] Santosh Kumar, et. ,al, "Role of Machine Learning in Managing Cloud Computing Security", Role of Machine Learning in Managing Cloud Computing Security", 78-1-6654-3789-9, IEEE, 2022.