IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

GESTURE CONTROLLED VIRTUAL MOUSE AND VOICE AUTOMATION WITH INTEGRATED GESTURE DATABASE

Revolutionizing Human-Computer Interaction with Seamless Gesture Recognition and Voice Command Integration

[1]H. S. Annapurna, [2]Koushik Umesh Pai, [3]Likhith Gowda M J, [4]Jitendra Patel N B, [5]Kushala H E

[1]Professor, [2]Student, [3]Student, [4]Student, [5]Student

Information Science and Engineering Department

Sri Siddhartha Institute of Technology, Tumkuru, India, 572105

Abstract: In recent years, human-computer interaction has advanced significantly, particularly with gesture control and voice automation technologies. This paper proposes a system integrating gesture control for virtual mouse manipulation and voice commands for automation tasks, enhancing user experience and productivity. Using computer vision techniques like hand tracking and gesture recognition, gesture control interprets hand movements as mouse actions, enabling contactless interaction with digital interfaces. Simultaneously, voice recognition and natural language processing (NLP) are employed to execute user commands, reducing reliance on manual input and making interactions seamless and efficient. The system includes a browser-based interface for managing and customizing gestures, with a database that allows users to store and manage their gestures. Upon executing the Python application, users can log in to access a dedicated page to view, add, or delete gestures, providing flexibility to tailor interaction methods to specific needs. This integration of gesture control and voice automation enhances accessibility for individuals with disabilities or limitations in traditional input methods. This innovative approach demonstrates the potential for multimodal input modalities to improve usability and efficiency in diverse computing environments, paving the way for more user-centric technology solutions.

Index Terms **-** Computer vision, Media Pipe, Real-time gesture recognition, Graphical User Interface, Hand Tracking.

I. Introduction

The generation that is now growing up using information technology (IT) tools, smartphones, display a different development of certain areas of their brains, which can mean not only biological, physiological, but also mental change. With the development of augmented-reality technology, researchers are working to reduce people's workload while increasing their productivity by studying human–computer interactions (HCI). The Natural User Interface (NUI) of hand-gesture recognition is an important topic in HCI. Hand-gesture-based interfaces allow humans to interact with a computer in the most natural way, typically by using fingertip movements [1]. Fingertip detection is broadly applied in practical applications, e.g., virtual mice, remote controls, sign-language recognition, or immersive gaming technology. Therefore, virtual mouse control by fingertip detection from images has been one of the main goals of vision-based technology in the last decades, especially with traditional red-green-blue (RGB) cameras.This system, designed using the latest technology, can recognize both static and dynamic gestures, along with voice commands, enhancing the naturalness and user-friendliness of the interaction. Implemented using

Convolutional Neural Networks (CNN) within the Media Pipe framework on pybind11, it requires no additional hardware and operates on Windows [2]. The Gesture Controlled Virtual Mouse comprises two modules: one using Media Pipe’s hand detection for tracking and interpreting hand movements into mouse actions like clicking, dragging, and scrolling, and another using uniformly colored gloves to improve recognition accuracy in varied lighting conditions. The first module benefits from Media Pipe's efficient, real-time hand landmark detection using CNNs, while the second employs color-based segmentation techniques to isolate hand movements from backgrounds, enhancing reliability [3].

Voice command integration adds versatility, enabling users to execute complex tasks through spoken instructions, powered by advanced voice recognition and natural language processing (NLP) technologies. This multimodal approach not only broadens interaction scope but also significantly improves user experience [4]. Additionally, the system includes a browser-based interface with a database where users can access predefined gestures and understand how they work, allowing for easy customization and management of gestures. Users can log in to view, add, or delete gestures, offering flexibility to tailor interaction methods to specific needs.The Gesture Controlled Virtual Mouse presents numerous advantages: it reduces physical strain associated with traditional input devices, offers a hygienic contactless alternative suitable for shared environments, and enhances accessibility for users with disabilities or motor impairments. In summary, the Gesture Controlled Virtual Mouse represents a major advancement in HCI by combining gesture recognition and voice automation for a natural, intuitive, and inclusive computer interaction [5]. This innovative system contributes to HCI research and sets the stage for further developments in multimodal input systems, emphasizing accessibility and user-centric design.

II. PROBLEM STATEMENT

TRADITIONAL INPUT METHODS IN HUMAN-COMPUTER INTERACTION (HCI), LIKE KEYBOARDS AND MICE, POSE LIMITATIONS SUCH AS PHYSICAL STRAIN AND LACK OF NATURAL INTERACTION. THE PROJECT AIMS TO DEVELOP A GESTURE CONTROLLED VIRTUAL MOUSE AND VOICE AUTOMATION SYSTEM TO ADDRESS THESE ISSUES. BY INTEGRATING ADVANCED MACHINE LEARNING AND COMPUTER VISION TECHNOLOGIES, THE SYSTEM INTERPRETS HAND GESTURES AND VOICE COMMANDS FOR INTUITIVE INTERACTION. ADDITIONALLY, A BROWSER-BASED APPLICATION WITH AN INTEGRATED GESTURE DATABASE ALLOWS USERS TO EASILY MANAGE AND CUSTOMIZE GESTURES, ENHANCING ACCESSIBILITY AND USER EXPERIENCE.

III. PROPOSED SYSTEM

A. OVERVIEW

The proposed system, "Gesture Controlled Virtual Mouse and Voice Automation with Integrated Gesture Database," offers a seamless and intuitive interaction experience. It comprises gesture control for virtual mouse manipulation and voice automation for task execution. Leveraging advanced machine learning and computer vision, the system accurately interprets hand gestures and voice commands. Additionally, an integrated gesture database allows users to manage and customize gestures easily through a browser-based interface. This system enhances accessibility, usability, and user experience in Human-Computer Interaction (HCI), paving the way for further advancements in multimodal input systems [5].

B. PALM DETECTION MODEL

The paper introduces a single shot detector model tailored for mobile real-time usage, focusing on hand identification and position detection using face mesh. It addresses challenges like wide range of hand sizes, occlusions, and lack of high contrast patterns in hand regions. The method utilizes a palm detector, non-maximum suppression, square anchor boxes, and a feature extractor based on a codec pair to enhance palm identification accuracy to 95.7%. Comparatively, a baseline accuracy of 86.22% is achieved without these enhancements. These techniques collectively improve hand detection in various scenarios, including social interactions and self-scenarios like handovers [5].

C. HAND LANDMARK MODEL

The Media pipe approach overcomes hand detection challenges by utilizing palm recognition followed by direct coordinate prediction. Their hand landmark model employs regression to pinpoint 2D and 3D hand knuckle positions accurately, even in partially obscured or incomplete hand images. Media pipe has annotated over 30,000 real-world photos with ground truth 3D coordinates and generated synthetic hand models to enhance coverage and accuracy. This comprehensive approach ensures consistent and robust hand posture representation, improving the overall performance of the system [6].

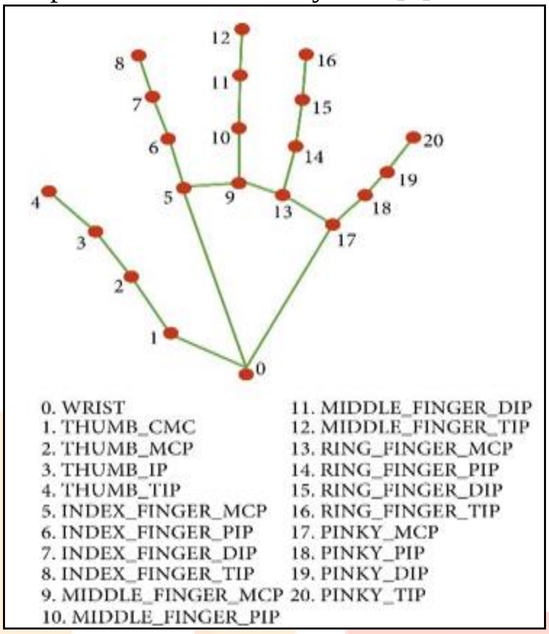


Fig.1 Landmarks in the hand

D. SYSTEM STRUCTURE FOR VOICE AUTOMATION

The proposed method for constructing an efficient Personal Voice Assistant utilizes a Speech Recognition library with various in-built capabilities. This allows the assistant to understand user commands and respond using Text-to-Speech operations. The system architecture includes a microphone to capture speech, transcription of audio files, and comparison of the input against pre-established rules to generate the expected outcome. This design ensures accurate interpretation and response to user instructions [4].

IV. METHODOLOGY

The suggested artificial intelligence virtual mouse system utilizes data from camera images on a personal computer or laptop. The Python computer vision package OpenCV is used to build a video capture object, which in turn triggers the web camera to begin recording. The frames are taken by a webcam and sent to a computerized artificial intelligence.

Setting major hand or minor hand using points from media pipe framework, obtaining maximum hands is 2, min detection confidence is 50% and min tracking confidence is 50 % and getting status of finger. Obtain current finger state 1 if fingers are open else 0. Set 'finger' by computing ratio of distance between fingertip, middle knuckle, base knuckle. Points for each finger [[8,5,0], [12,9,0], [16,13,0], [20,17,0]] from index to pinky for thumb is zero [6].

Using the media pipe package of Python, if the index finger and the middle finger is raised, then mouse cursor is caused to move around the window of the computer. The detected and tracked hands are collected here, with each hand being represented by a list of 21 landmarks. The x and y coordinates are scaled to [0.0, 1.0] and (x, y, and z coordinates). It is based on the image's width and height. With the wrist as the starting point, and the lower the value, the depth of a landmark is represented by z and the closer the landmark is to the camera. Z's magnitude is about comparable to that of x.

A trove of detected/tracked hands, with each one represented by a list of 21 landmarks in global coordinates. The x, y, and z components of each landmark are real-world three-dimensional coordinates in meters, with the hand's approximate geometric center serving as the point of origin. Handedness data from identified and followed hands (i.e. is it a left or right hand). An individual's score and label make up their hand. Label can have the values "Left" or "Right" in a string. Score is the expected handedness probability, and it is greater than or equal to 0.5 in every case [5].

- If V shape gesture is detected with index and mid finger then mouse movement from obtaining absolute vale from x and y axis. If Fist gesture recognition, then drag and drop and multiple select function will be handled.
- If Pinch gesture is detected through major hand, on x-axis Control change system brightness in fig 5, on y axis Control change system volume in fig 6.
- If pinch gesture is detected through Minor hand, on x-axis 'Control scroll Horizontal', on y-axis 'Control scroll Vertical'. Obtaining pinch level with distance and position and assign threshold.
- If mid finger is open with 4 fingertip id, then left click, Index finger is open with fingertip id then right click, two finger closed then double click.

All function can be trigger and execute through pyautogui which helps to interact with os and execute mouse functions.

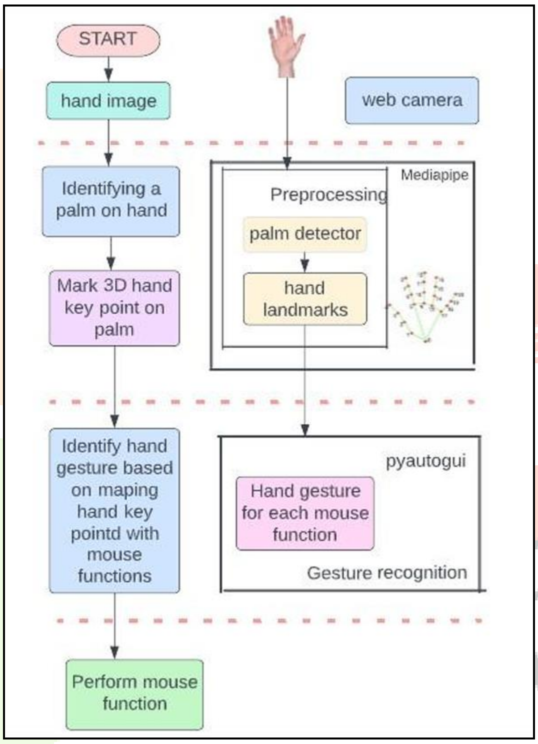


Fig.2 Flow chart of Gesture Controlled Virtual Mouse System

Methodology used by voice automation Headings, From the get-go, we utilize sapi5 and pyttsx3 to provide our software the ability to interact with the system voice. The pyttsx3 library is a Python implementation of text-to-speech technology. It's compatible with Python 2 and 3, unlike many other libraries, and it even functions while you're not online. Windows programmed may take use of voice detection and synthesis thanks to the Speech Application Programming Interface (SAPI), an API created by Microsoft [6]. The program's capabilities are then defined in the main function. The suggested system is expected to be able to do the following.

a) The helper constantly polls the user for feedback and awaits further instructions. The listening time may be adjusted based on the user's needs.

b) The helper will keep asking the user to repeat the instruction if it doesn't understand it the first time around.

c) Depending on the user's preferences, this assistant can be programmed to speak in a male or female voice.

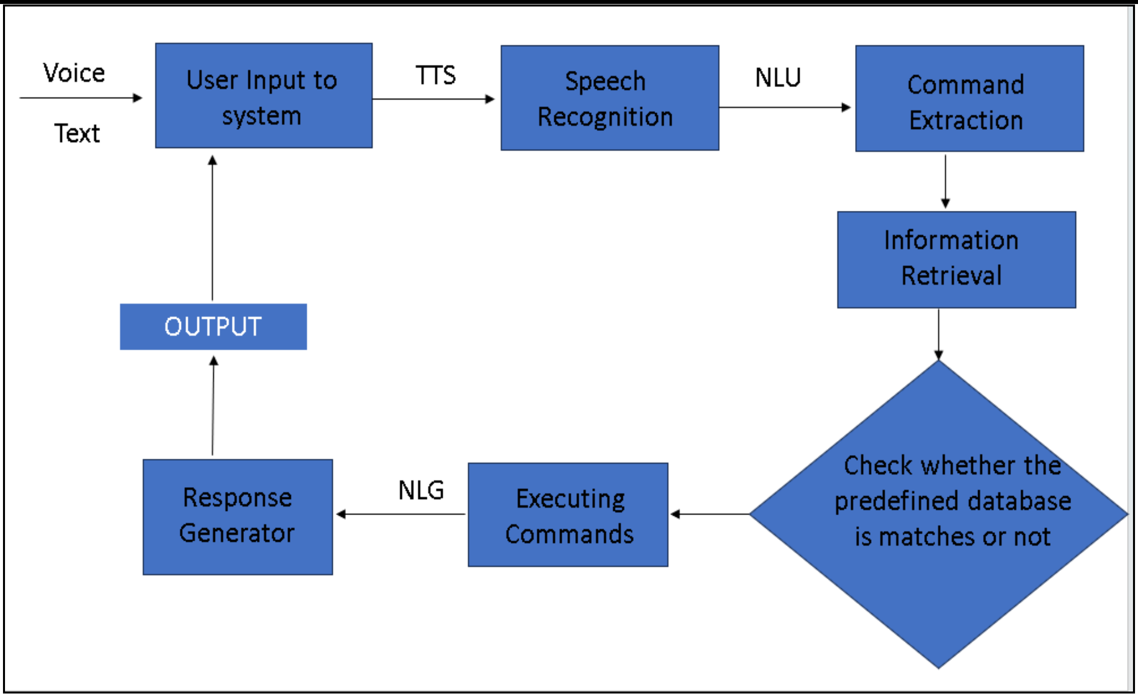
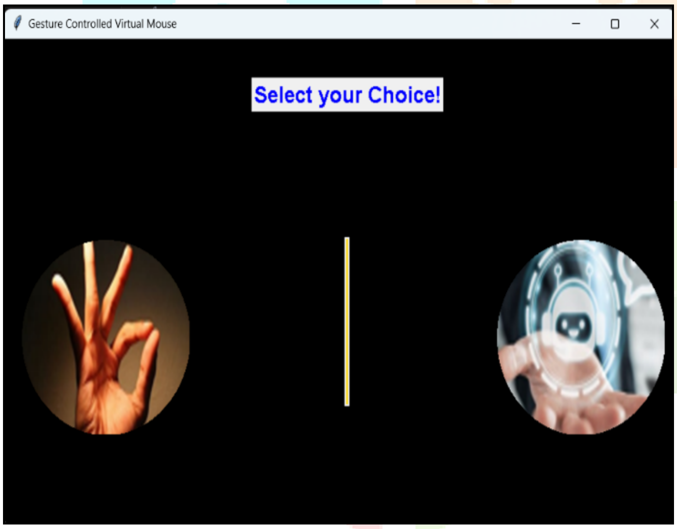


Fig.3 Flowchart of Voice Automation Model

V. RESULT ANALYSIS



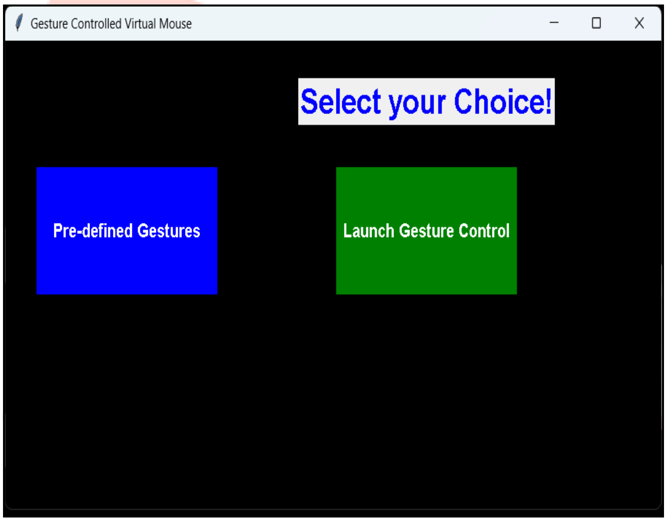


Fig.4 Snapshot of User Interface

Fig.5 Snapshot of User Interface when click on Gesture Control

When the program runs, it displays a user interface featuring two image buttons: one for a gesture-controlled virtual mouse and the other for voice automation. Clicking on the gesture-controlled virtual mouse button opens another user interface that provides predefined gestures and launches the gesture control system, allowing users to operate the mouse and other systems [7].

When the user launches the gesture control, the system will access the camera and perform tasks such as:

a) Neutral Gesture: Neutral Gesture. Used to halt/stop execution of current gesture.

b) Move Cursor: Cursor is assigned to the midpoint of index and middle fingertips. This gesture moves the cursor to the desired location. Speed of the cursor movement is proportional to the speed of and.

c) Left Click: Gesture for single left click.

d) Right Click: Gesture for single right click.

e) Double Click: Gesture for double click.

f) Scrolling: Dynamic Gestures for horizontal and vertical scroll. The speed of scroll is proportional to the distance moved by pinch gesture from start point. Vertical and Horizontal scrolls are controlled by vertical and horizontal pinch movements respectively.

g) Drag and Drop: Gesture for drag and drop function ability. Can be used to move/transfer files from one directory to other.

h) Multiple Item Selection: Gesture to select multiple items.

i) Volume Control: Dynamic Gestures for Volume control. The rate of increase/decrease of volume is proportional to the distance moved by pinch gesture from start point.

j) Brightness Control: Dynamic Gestures for Brightness control. The rate of increase/decrease of brightness is proportional to the distance moved by pinch gesture from start point.

Fig.6 Snapshot of Neutral Gesture

Fig.7 Snapshot of V- Gesture

Fig.8 Snapshot of Pinch Gesture

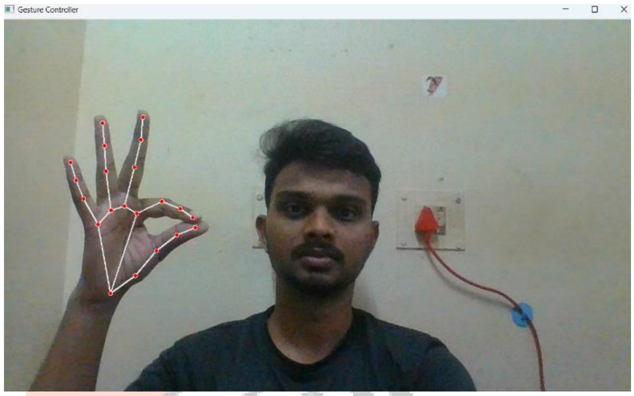


Fig.9 Snapshot of Opposite pinch Gesture

The project incorporates a voice assistant to enhance functionality, enabling users to:

- Launch or stop gesture recognition
- Perform Google searches and click on the links
- Find locations on Google Maps
- Navigate files
- Check date and time
- Execute copy and paste commands
- Sleep or wake the voice assistant
- Exit the application

These features make the system versatile and user-friendly.

When first activated, the assistant waits for user instructions. It records the user's voice command, searches for specified keywords, and if a relevant keyword is identified, executes the requested action. Results are provided verbally and in text on the terminal. If the input is improper, the assistant will wait for a correct input. These features ensure the smooth and versatile operation of the system [7].

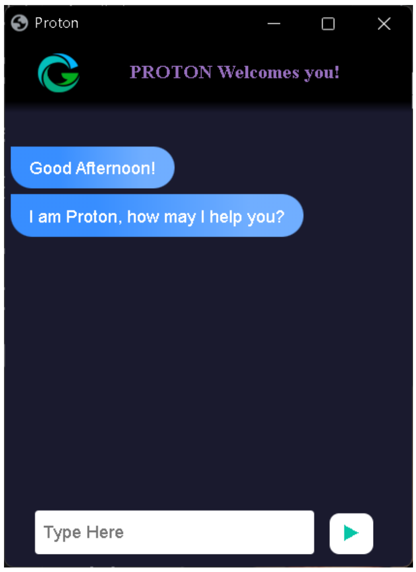


Fig.10 Snapshot of User Interface when click on Voice Automation

VI. CONCLUSION

The "Gesture Controlled Virtual Mouse and Voice Automation with Integrated Gesture Database" project represents a significant advancement in Human-Computer Interaction (HCI). By integrating gesture recognition and voice automation, the system offers an intuitive, contactless, and user-friendly interaction experience. The use of advanced machine learning, computer vision techniques, and natural language processing (NLP) ensures accurate and efficient interpretation of user commands. The inclusion of a browser-based interface with an integrated gesture database allows users to manage and customize gestures easily, enhancing the system's flexibility and adaptability to individual needs. This project not only addresses the limitations of traditional input methods but also provides a hygienic and accessible alternative, particularly benefiting users with disabilities or motor impairments. Overall, this innovative system demonstrates the potential of multimodal input technologies to improve usability and accessibility in diverse computing environments. It sets the stage for future developments in HCI, emphasizing user-centric design and the importance of seamless interaction methods.

VII. REFERENCE

[1] J. Katona, “A review of human–computer interaction and virtual reality research fields in cognitive Info Communications,” *Applied Sciences*, vol. 11, no. 6, p. 2646, 2021.

[2] D.-S. Tran, N.-H. Ho, H.-J. Yang, S.-H. Kim, and G. S. Lee, “Real-time virtual mouse system using RGB-D images and fingertip detection,” *Multimedia Tools and Applications Multimedia Tools and Applications*, vol. 80, no. 7, pp. 10473–10490, 2021.

[3] D.-H. Liou, D. Lee, and C.-C. Hsieh, “A real time hand gesture recognition system using motion history image,” in *Proceedings of the 2010 2nd International Conference on Signal Processing Systems*, IEEE, Dalian, China, July 2010.

[4] Pawan R Zadgonkar, Abhishek R Waghate, Sakshi S Nivalkar, Pooja A Kondvilkar, Mrunmayee Hatiskar, “Vision: The Desktop Voice Assistant”, ISSN: 2320-2882, Vol. 12, Issue 4, April 2024.

[5] Prithvi J, S Shree Lakshmi, Suraj Nair and Sohan R Kumar, “Gesture Controlled Virtual Mouse with Voice Automation”, ISSN: 2278-0181, Vol. 12 Issue 04, April-2023.

[6] Likitha R, Kumaraswamy S, Revathi B, “AI Virtual Mouse System Using Hand Gestures and Voice Assistant”, published in “International Journal of Engineering Research and Applications” ISSN: 2248-9622, Vol. 12, Issue 12, December 2022.

[7] Dubbaka Megha Sai Reddy, Seilekha Kukkamudi, Rishika Kunda, Mamatha, “Virtual Mouse using Hand Gesture”, International Conference on Knowlegde Engineering and Communication System (ICKECS), December 2022.

[8] Cherukupally Karunakar Reddy, Suraj Janjirala, Kevulothu Bhanu Prakash, “Gesture Controlled Virtual Mouse with the Support of Voice Assistant”, published in “Ijraset Journal For Research in Applied Science and Engineering Technology” ISSN:2321-9653, Jun 2022