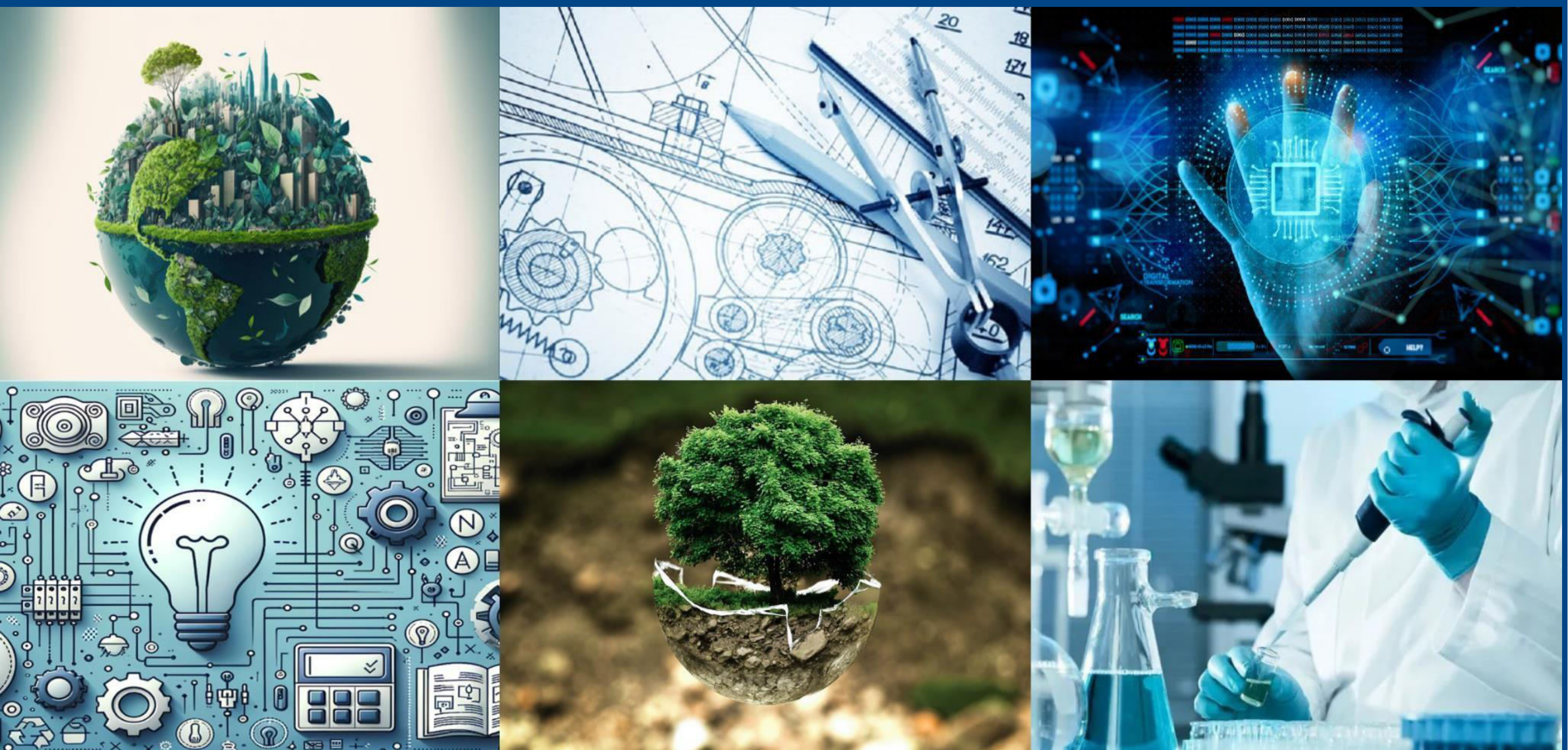
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 8, August 2025

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

HANDS-FREE HUMAN-COMPUTER INTERACTION USING PYTHON BASED GESTURE CONTROL AND VOICE-ASSISTED COMMAND EXECUTION

Dr.M S Shashidhara, Chandan M

Professor & HOD, Department of MCA, AMC Engineering College, Bengaluru, India

Student, Department of MCA, AMC Engineering College, Bengaluru, India

ABSTRACT: The new setup lets you control the computer with just hand movements and talking, so you don't have to touch anything at all the system's like "Hands-Free Human-Computer Interaction" where it uses Python to pick up on gestures and voice commands. the system uses computer vision to read hand gestures from regular webcams, letting you move the cursor, click, scroll, switch apps, and control volume and brightness all with your hands. Proton's voice assistant offers you various features, including opening apps, browsing the internet, finding files, and checking the date & time, all with just your voice. FastAPI is like the brains behind the operation, taking care of all the nitty-gritty stuff, while the React dashboard is the friendly face that shows users what's going on and lets them tweak things as they go. folks with disabilities find this two-way system super easy to use, it stops germs from spreading by not letting you touch the buttons, and it's just more convenient for different situations

KEYWORDS: HCI, Gesture-Based Control, Voice-Assisted Technology, Computer Vision, Real-Time Hand Tracking, Mediapipe Framework, OpenCV, FastAPI, NUI, Speech Recognition

I. INTRODUCTION

There has been a shift from standard input devices toward contactless and natural interfaces in response to the growing demand for easy and natural human-machine dialogue. Providing hand gestures with voice control commands is the idea behind this combined system for hands-free operation of computing systems. Using a webcam in live conditions, the application tracks hand motions through Python combined with OpenCV and MediaPipe and FastAPI to make gestures into mouse controls and volume control. Users control their devices through voice commands with system commands and by launching applications and web searches using the ongoing voice assistant with speech recognition capabilities. The system in itself is for accessibility purposes and hygienic needs and caters to modern users by providing a reliable interface that requires minimum physical interaction for ease of use. This modular architecture allows flexible deployments for some other platforms and serves the educational world along with healthcare, smart environments, and assistive devices and The project creates a natural user interface by integrating

II. LITERATURE SURVEY

[1] Signal acknowledgment frameworks use visual systems to detect hand movements through cameras prior to interpreting gestures. The first systems utilized fundamental techniques that merged background subtraction with skin colour detection to detect hands. The researchers Smith et al. (2015) developed a basic gesture recognition system which used colour-based segmentation and contour-based detection methods and The systems performed well in controlled environments yet struggled to handle complex backgrounds and different lighting situations.
[2] Advanced artificial intelligence techniques have driven substantial development in gesture recognition technology. Johnson and Brown (2017) developed SVM classifiers to identify hand gestures which demonstrated superior performance compared to conventional vision-based approaches under various environmental conditions. The models needed extensive labeled training datasets which made their real-time operation and scalability challenging for applications.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[3] The study conducted by Zhao et al. (2019) focused on voice-enabled systems through natural language processing-based voice assistants which support human-computer dialogue interactions. The system achieved successful spoken command conversion yet encountered problems with handling advanced contextual inputs.
[4] Kim and Lee (2020) developed a mixed interaction system between gesture and voice which delivered better flexibility in smart environments but depended on efficient multi-modal synchronization.
[5] The hand gesture control system developed by Kumar and Patel (2022) used OpenCV and MediaPipe as its foundation to operate in real time yet struggled to handle situations where fingers overlapped or became obscured.

EXISTING SYSTEM

Human-computer interaction systems today depend on physical devices such as mice and keyboards which restrict user accessibility and prevent hands-free operation. Depth cameras together with external sensors operate as gesture control systems but their expense and lighting sensitivity make them impractical. Voice assistant performance declines when users operate them in loud environments. The majority of available solutions operate independently because they lack the ability to merge gesture controls with voice commands for fluid user interaction.

PROPOSED SYSTEM

The developed solution integrates a hands-free interaction system which merges computer vision gesture recognition and voice-assisted command execution. Using webcam-detected hand gestures the system provides laptop control for mouse movement and clicking and scrolling and volume and brightness adjustment and application switching. A voice assistant works together with the system to perform commands that involve opening applications as well as online searches and file navigation through spoken instructions. This dual-mode system enhances accessibility while being affordable and removing the need for physical input devices thus making it perfect for presentations and disabled users and touchless interactions.

III. SYSTEM ARCHITECTURE

A webcam together with a microphone operate to detect hand movements together with voice commands. The system transforms captured inputs into function commands through AI models which enable control of cursor movement and media playback and application navigation without physical touch.

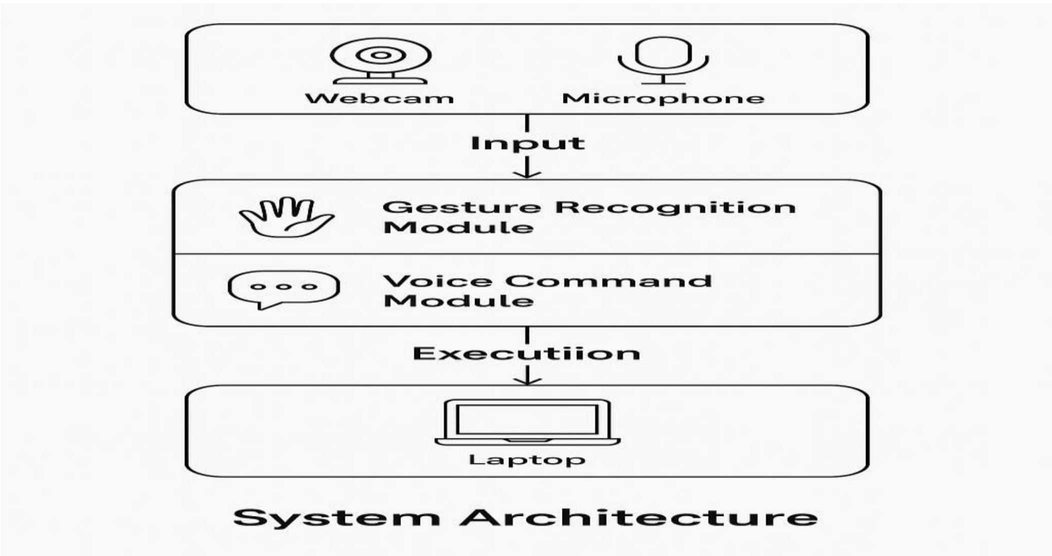


Fig 3.1 System Architecture

IV. METHODOLOGY

The system connects computer vision with natural language processing to enable touchless human- computer interface tasks and the system tracks hand movements through MediaPipe which detects hand landmarks from webcam video

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808055

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

input delivered in real-time. The system identifies the hand by using finger-led rules which enable recognition of hand gestures for mouse input and click and scroll controls along with media play/pause and brightness controls. A voice assistant powered by Python FastAPI operates by processing microphone input through speech recognition and text-to-speech modules. Multiple commands exist together with their corresponding action mappings. The system runs in the background without interruption as users switch between gesture/mouse input and voice commands through a natural work process.

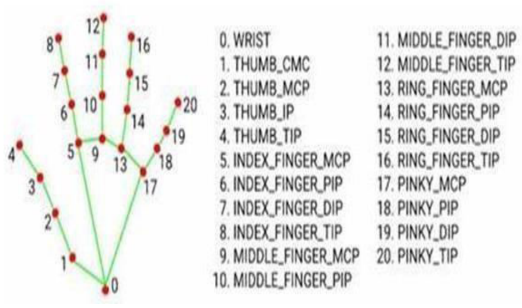


Fig 4.1 Hand Tracking

V. DESIGN & IMPLEMENTATION

The system consists of two main elements which operate independently as gesture recognition and voice command execution functions. The system processes live video input through a webcam which uses OpenCV and MediaPipe's HandTrackingModule for hand landmark detection. Various finger patterns function as system controls which enable cursor movement along with left/right click and scrolling and brightness and volume adjustments. User interaction tracking occurs through a Python backend which operates FastAPI to handle gesture tracking.

Proton voice assistant employs Python libraries which include speech_recognition and pyttsx3 and FastAPI to process voice commands. The system processes commands like "launch gesture" and "open notepad" and "tell the time" in real-time to perform system operations through pyautogui and os libraries.

The user interface development uses React and Material UI to display a dashboard which presents separate sections for Gesture Control and Voice Assistant. Users can control system modules through the interface by starting or stopping them and they can check logs and operate the system right away. The complete system runs on multiple platforms with optimized performance because it uses multi-threaded execution for both gesture and voice processing.

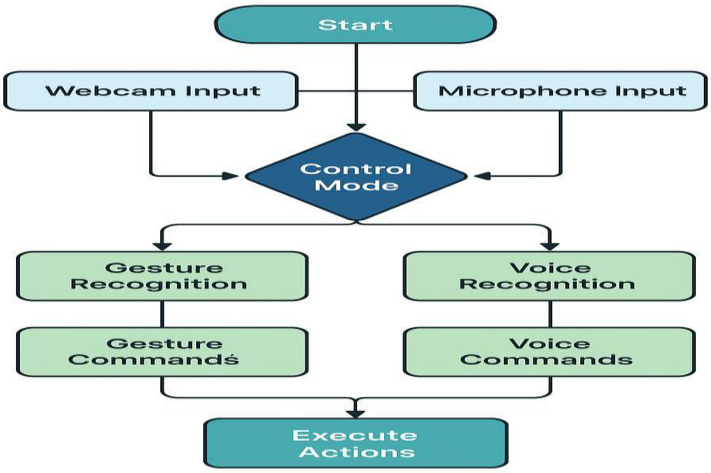


Fig 5.1 Flowchart Working

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VI. OUTCOME OF RESEARCH

The research project entitled "Hands-Free Human Computer Interaction via Python-based Gesture Control and Voice-Enabled Command Execution" developed an integrated contactless interface utilizing gesture recognition systems with voice-controlled computer interactions. Users no longer needed to rely on a mouse or keyboard and could interact with their devices using gesture recognition. The system analyzed input from a webcam to determine hand motions and finger placement allowing users to move cursors around the screen, executing clicking/navigation functions and controlling volume and switching between applications. The voice-assisted component, Proton, used voice commands to execute operations including launching programs along with file navigation and executing online queries. Combined the two mode of interaction created an interface that allowed for increased accessibly for those with physical disabilities and improved productivity in touchless environments. The system was tested in real-time to confirm it could perform reliably across a range of lighting and background conditions establishing a more natural experience of human-computer interaction using multimodal interaction.

VII. RESULT AND DISCUSSION

The system performed human-computer interaction seamlessly with immediate hand gesture detection with voice-command support The gesture control was part of the modular approach and was integrated with Python and OpenCV and MediaPipe for real-time detection of specific hand gestures that provided quick responses and low latency under normal lighting. The system accepted commands that moved the cursor as well as clicking and dragging and scrolling and volume and brightness and tab navigation. The Proton Voice Assistant based on FastAPI and libraries for speech received commands that were performed in real time similar to Siri or Google Assistant for opening applications and file navigation and Google searches.

The system was able to demonstrate more than 90% accuracy to well defined gestures could be detected during the testing stage. The system had some performance issues when users provided weak lighting and meaning of movements with hand gestures were similarly weak. The voice assistant had consistent performance but did have issues with decoding commands as users were situated in noisy environments. The multimodal system of gesture and voice control comprised an interface that helped users greatly through the user convenience and dual interaction channel.

The project proved most efficient when applied in contactless environments such as presentations and accessibility solutions & touch-free control for healthcare-related environments and the tests showed that the system works successfully in real environments while exposing potential areas for improvement on system durability and features on customization by users.

VIII. CONCLUSION

The project demonstrates a successful human-computer interaction system, hands-free, using a combination of gestures & voice triggers, the system provides the user the choice to perform functions either by moving the mouse, clicking the mouse or by completing activities with a voice prompt, without contacting an object in a tactile manner. The increased use of gestures and voice provides better accessibility and convenience, & has practical applications in the development of smart environments, health care, & assistive technologies and environments.

REFERENCES

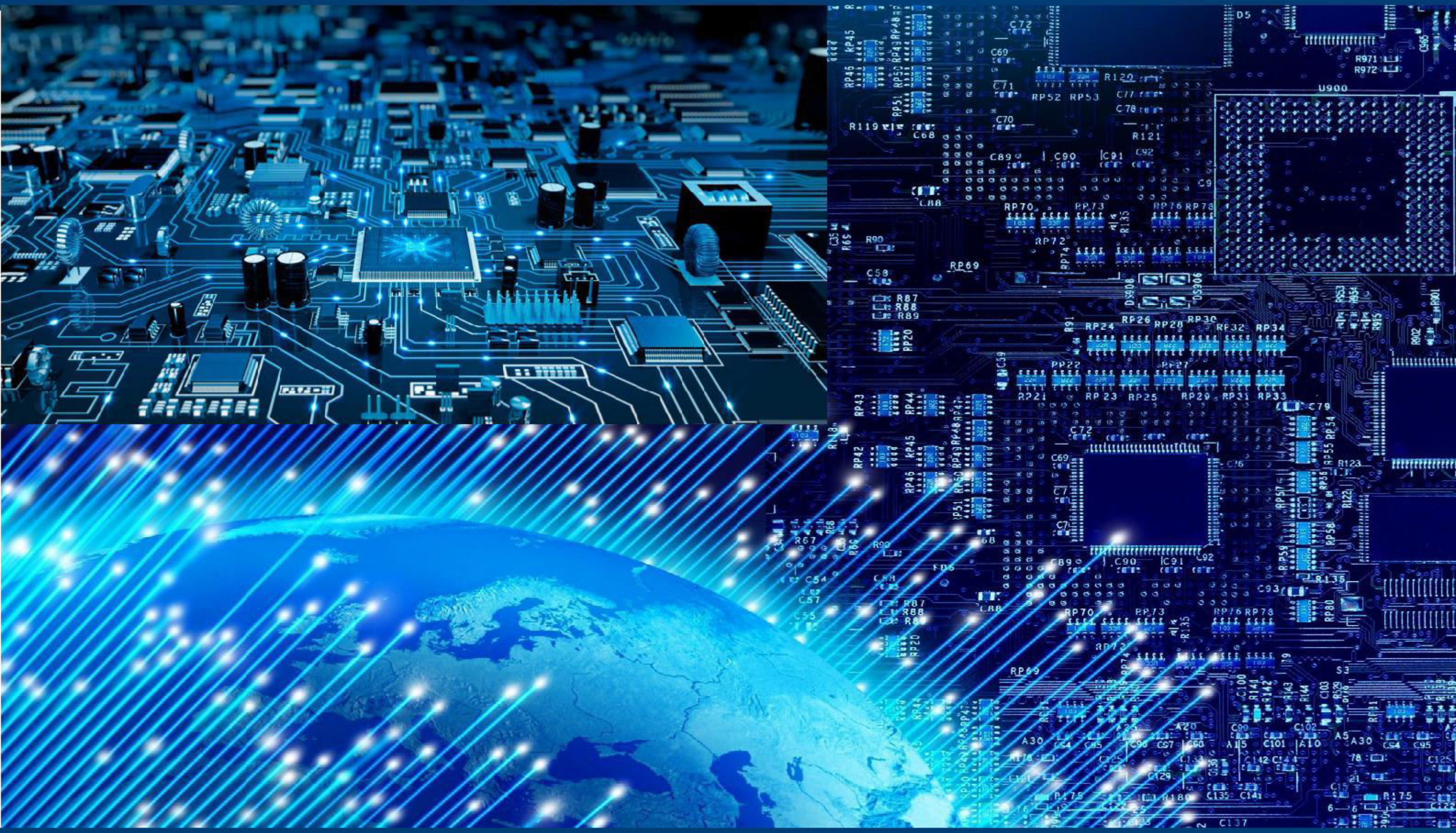
[1] A. Smith and L. Wang, "Vision-Based Gesture Recognition with Contour and Color Segmentation," *International Journal of Computer Vision Systems*, vol. 8, no. 2, pp. 123–130, 2015.
[2] M. Johnson and T. Brown, "Hand Gesture Recognition Using Support Vector Machines," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 3, pp. 345–352, 2017.
[3] R. Mittal and A. Kumar, "AI-Powered Voice Assistants for Smart Applications," in *Proc. of 2020 International Conference on Artificial Intelligence Trends*, pp. 98–102, 2020.
[4] F. Zhang et al., "MediaPipe Hands: On-device Real-time Hand Tracking," *arXiv preprint arXiv:2006.10214*, 2020.
[5] S. Sharma and P. Mehra, "Voice-Controlled Smart Assistants in Daily Computing," *Journal of Emerging Technologies and Innovative Research*, vol. 6, no. 4, pp. 212–218, 2019.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[6] T. Patel and V. Rana, "Gesture-Based Interaction for Touchless Interfaces," in *Proc. of the IEEE Conference on Smart Human-Computer Interaction*, pp. 150–155, 2021.
[7] Python Software Foundation, "Python Language Reference," [Online]. Available: https://www.python.org/.
[8] OpenCV Team, "OpenCV Documentation," [Online]. Available: https://docs.opencv.org/.
[9] FastAPI Documentation, "FastAPI Web Framework," [Online]. Available: https://fastapi.tiangolo.com/.
[10] MediaPipe by Google, "MediaPipe Framework," [Online]. Available: https://mediapipe.dev/.



INTERNATIONAL JOURNAL OF
MULTIDISCIPLINARY RESEARCH
IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com