Virtual Mouse & Voice Assistance

Akshay Godhale
Department of Information Technology
Vidyalankar Institute of Technology, Mumbai, India
akshay.godhale@vit.edu.in

Ayush Warang
Department of Information Technology
Vidyalankar Institute of Technology, Mumbai, India
ayush.warang@vit.edu.in

Vedant Kajrekar
Department of Information Technology
Vidyalankar Institute of Technology, Mumbai, India
vedant.kajrekar@vit.edu.in

Rohit Barve
Department of Information Technology
Vidyalankar Institute of Technology, Mumbai, India
rohit.barve@vit.edu.in

***Abstract*— Artificial intelligence technology has paved the way for hand gesture recognition in controlling virtual devices. This paper introduces a system that uses AI to identify hand gestures and convert them into movements of a virtual mouse, providing an alternative interface for those who find traditional mice or keyboards challenging. The system captures images of the user's hand using a camera, which are then processed by an AI algorithm to identify the gestures. These gestures are translated into corresponding mouse actions on a virtual screen. The system is adaptable and scalable, capable of being controlled virtually using dynamic/static hand gestures and a voice assistant. It utilizes Machine Learning and Computer Vision algorithms without any additional hardware requirements. The model is implemented using a Convolutional Neural Network and the mediapipe framework. This system could potentially enable hands-free operation of devices in hazardous environments and provide an alternative to hardware mice, thereby enhancing user experience and accessibility**

Keywords*— *Computer vision, hand gesture recognition, Media-pipe, voice assistant & virtual mouse.

I. INTRODUCTION

The contemporary world is inundated with technological advancements that permeate our daily lives. Among these technologies, computer technologies continue to burgeon worldwide, facilitating tasks beyond human capability. The interaction between humans and computers often hinges on output devices like the mouse, essential for navigating graphical user interfaces (GUIs) through pointing, scrolling, and movement. However, traditional hardware mice and touchpads are not always conducive to complex tasks and can be cumbersome to carry around.

Over time, mouse functionality has transitioned from wired to wireless, enhancing ease of movement and functionality. Concurrently, speech recognition technology has emerged, enabling voice-controlled functions for search and translation purposes, albeit with some latency in performing mouse functions. Subsequent advancements in human-computer interaction introduced eye-tracking techniques for cursor control, albeit with limitations for users wearing contact lenses or having long eyelashes.

Developers have explored various models for human gesture recognition, often requiring expensive gloves, sensors, and color caps for fingertip tracking. Amidst these developments, artificial intelligence (AI) stands out, revolutionizing myriad sectors and offering solutions to existing challenges through cutting-edge algorithms and tools.

One such innovative application is the hand gesture-controlled virtual mouse, empowered by AI. This technology enables users to manipulate their computer mouse using hand gestures, obviating the need for a physical mouse. By leveraging camera vision-based approaches, the system tracks the user's hand movements and translates them into mouse functions. Computer vision algorithms analyze video input to identify and track hand movements, while machine learning models recognize specific gestures and execute corresponding mouse actions.

This technology presents numerous advantages, notably enhancing accessibility and providing a more intuitive user experience. It offers a natural and intuitive means of interaction, particularly in scenarios where physical mice or touchpads are impractical. Its versatility extends to gaming, virtual reality, and accessibility, promising to make computing more seamless and inclusive for all users.

II. LITERATURE REVIEW

Research in the realm of AI-driven virtual mouse systems has seen various approaches. Initially, glove-based methods were employed for data collection, while later systems utilized colored paper attached to hands for gesture recognition. However, these methods often lacked accuracy in mouse operations. Glove-based approaches posed issues such as impracticality, potential allergies for users with sensitive skin, discomfort during prolonged use leading to sweating and skin reactions. Similarly, colored paper tips for gesture recognition were not consistently effective.

Recent contributions have leveraged Google's work with the mediapipe framework to develop gesture-controlled virtual mouse systems. These systems enable users to manipulate the mouse cursor and perform functions like left click, right click, drag and drop, volume control, and brightness control through hand gestures. Efforts have also focused on camera-based hand gesture recognition interfaces.

[1]. This analysis discusses the development of a hardware-based system. While this model achieves remarkable accuracy, executing many movements proves challenging while wearing a glove, which significantly restricts the user's hand's range of

motion, speed, and agility. Prolonged glove usage can also lead to skin issues, making it less suitable for users with sensitive skin types.[2].A machine-user interface was developed utilizing straightforward computer vision and multimedia techniques for hand gesture detection. However, a notable drawback is that skin pixel identification and hand segmentation from stored frames are prerequisites before implementing gesture comparison techniques.[3].In this study, they presented a system for recognizing hand movements which utilizes a mobile phone's camera and a connected mobile projector as a visual feedback medium. Their framework allows other mobile applications to seamlessly integrate for gesture recognition. This architecture facilitates the swift and straightforward creation of research prototypes supporting gestures, thereby shifting the user's focus from the device to the content.[5]. This study delves into the advanced exploration of robots with gesture controls. The initial section provides insights into the art of hand gesture identification concerning how they are perceived and captured by conventional video cameras. Motion features are extracted based on estimations of smoothed optical flow. Additionally, face detection is employed to generate a user-centric representation of this data, followed by training an effective classifier for differentiation purposes.[6].In this model, the hand's center is identified, and the hand's calculated radius is determined. By employing the convex hull technique, fingertip points are established. These hand gestures are utilized to govern every mouse movement. However, a notable challenge with this approach is that the frame must be initially saved before undergoing detection, resulting in longer processing times than what is required for real-time applications.[8].The system has the capability to generate colored masks employing techniques for color variation. Subsequently, mouse functions are executed through hand gestures. However, the implementation of this approach is challenging.[9].The study aimed to develop a virtual mouse system controlled by hand gestures, enhancing human-computer interaction. It explored technical aspects and algorithms for precise gesture recognition and mouse control, proposing innovative solutions. The authors discussed existing literature on gesture recognition, interface design, and human-computer interaction to contextualize their work, drawing insights from related studies on virtual input devices and user interface technologies.

III. PROPOSED SYSTEM

The proposed system incorporates a high-quality finger and hand tracking system known as MediaPipe Hands, which leverages machine learning (ML) to discern 2D and 3D landmarks of a hand from a single image. Unlike current state-of-the-art methods that often necessitate robust PC settings for inference, our solution offers real-time performance on mobile phones, accommodating multiple hands simultaneously. We hope that by making these hand perception capabilities accessible to the broader research and development community, it will inspire the creation of novel use cases and the exploration of new research avenues.

A. Palm Detection Model

We have developed a single-shot detector model tailored for mobile real-time usage to identify hand positions using face mesh. The MediaPipe Lite model and the full model are designed to recognize hands across a wide range of sizes and scales, even in occluded and self-occluded states, posing a challenging problem due to the absence of high-contrast patterns in the hand region. However, precise hand localization is facilitated by additional information such as torso and arm cues or human traits.

Our approach is interdisciplinary, employing various techniques to address these challenges. We first train a palm detector instead of a hand detector, as it is simpler to estimate the bounding boxes of rigid objects like palms and fists. The non-maximum suppression method proves particularly useful in social and self-scenarios due to the small size of the palms involved. Furthermore, modeling palms using square anchor boxes and employing a feature extractor based on a codec pair aid in grasping the entire scene context. To mitigate attentional drift during training, the palm model is designed to maintain a large number of hooks.

Through the combination of these methods, the MediaPipe model achieves an average palm identification accuracy of 95.7%. To provide context, a baseline accuracy of just 86.22% is achieved when using a standard cross-entropy loss and no decoder.

B. Hand Landmark Model

The Mediapipe approach addresses these challenges by employing a variety of techniques. Once the palm is recognized across the entire image, or through direct coordinate prediction, the Mediapipe hand landmark model utilizes regression to precisely pinpoint the 2D and 3D positions of hand knuckles within the observed hand areas. This model enables the development of a consistent internal hand posture representation, remaining resilient even when only a portion of the hand is visible or when the hand is partially obscured by the model's own body.

Furthermore, Mediapipe has meticulously labeled over 30,000 real-world photos with 21 3D coordinates to serve as ground truth data. Additionally, Mediapipe renders a high-quality synthetic hand model across various backgrounds and aligns it with the associated 3D coordinates to comprehensively cover available hand positions, providing valuable insights into hand geometry.

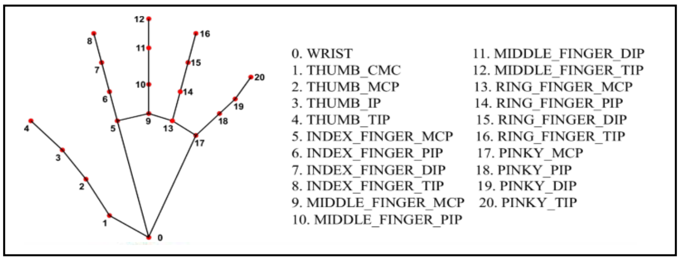


Fig. 1. Hand Landmark Model.

C. System Structure For Voice Assistant

The proposed system architecture for the Voice Assistant involves the utilization of a Speech Recognition library with built-in capabilities. This enables the assistant to understand user commands and respond through Text-to-Speech operations. Upon recording a user's voice instruction, the system employs speech-to-text techniques, including:

1) Conceptual Design: The overall design framework of the voice assistant system, outlining its components and functionalities.

2) Microphone Integration: Incorporating a microphone component for capturing speech patterns and user commands.

3) Audio Transcription: Transcribing audio files into text, which involves comparing the input against a predefined set of rules.

4) Outcome Generation: Processing the transcribed input to generate the expected outcome or response from the voice assistant.

By following this structured approach, the voice assistant system can effectively interpret user instructions and provide appropriate responses in a conversational manner.

D. Simple Procedure

The primary procedure of a voice assistant involves the conversion of spoken words into written text, a process known as "speech recognition." Subsequently, the computer utilizes the character set of the command to locate and execute the relevant script. However, there are additional complexities to consider. Despite efforts, background noise significantly impacts the effectiveness of voice recognition equipment. This challenge arises from the difficulty people face in distinguishing between speech and background noises such as a dog barking or a helicopter passing overhead. Addressing this issue is crucial for optimizing the performance of voice recognition system

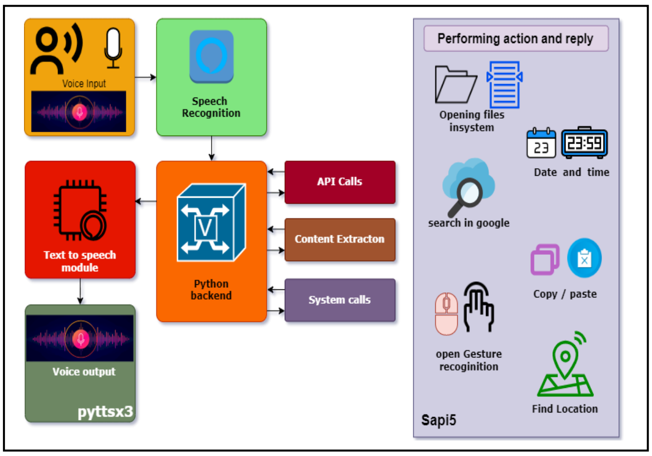


Fig. 2. Basic Work flow of voice assistant.

E. Process Flow

Voice assistants like Siri, Google Voice, and Bixby have become integrated into our mobile devices, while sales of smart speakers such as the Amazon Echo and Google Home are rising rapidly, mirroring the growth of smartphone sales a decade ago, as reported by recent NPR research. Approximately one in six Americans currently owns a smart speaker. Despite this progress, the integration of voice technology in the workplace still faces challenges, particularly in open office environments where excessive noise can be disruptive.

The process flow of a voice assistant typically involves three main stages:

1) Voice Recognition: The assistant's ability to recognize and interpret the user's voice commands.

2) Input Processing: Analyzing the user's input to determine its meaning and context, and applying appropriate actions.

3) Audio Transcription: Transcribing audio files into text, which involves comparing the input against a predefined set of rules.

4) Real-time Response: Providing the user with timely and relevant outcomes or responses in voice format.

The assistant initiates by gathering data from the user, capturing the user's spoken input and converting it into digital text for analysis and processing.

This structured approach ensures that voice assistants can effectively understand user commands, process them accurately, and deliver appropriate responses in real-time, thereby enhancing user interaction and productivity.

IV. Result and Implementation

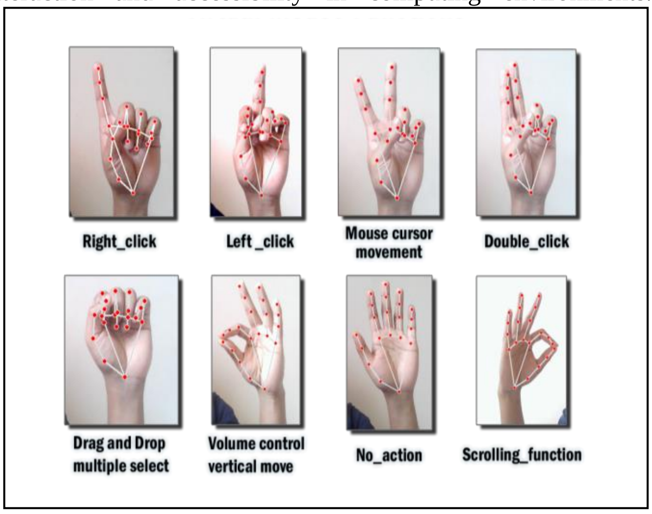
In the proposed AI virtual mouse system, the focus lies on enhancing human-computer interaction through computer vision. However, evaluating the AI virtual mouse system poses challenges due to the limited availability of datasets. Various tests have been conducted by positioning the webcam at different distances from the user to track hand gestures and detect hand tips under diverse illumination conditions.

The AI virtual mouse technology has been rigorously tested in different lighting environments, ranging from bright to dim, and varying distances from the webcam, including close proximity and four feet or more away from the screen. While the overall accuracy of the AI virtual mouse technology reached 99 percent, the precision of the right-click gesture remained lower due to its inherent complexity, presenting challenges in achieving precise mouse actions.

Nevertheless, all other gestures demonstrated superb accuracy, showcasing significant advancements compared to previous techniques. The proposed AI virtual mouse model exhibited exceptional performance, surpassing existing virtual mouse models in terms of accuracy. Notably, its novel feature lies in its capability to virtually manage computer functions akin to a physical mouse, encompassing tasks such as left and right clicks, scrolling, and mouse movements.

Moving forward, further refinement and optimization of the AI virtual mouse system are essential to address challenges associated with gesture recognition and improve the precision of complex actions like right-clicking. Additionally, expanding the dataset and conducting comprehensive testing across diverse user scenarios will enhance the robustness and reliability of the virtual mouse technology.

The proposed model's ability to emulate physical mouse functionalities opens up new avenues for enhancing user interaction and accessibility in computing environments.



Continued research and development efforts in AI virtual mouse technology hold promise for revolutionizing human-computer interaction and advancing the capabilities of virtual input devices in various applications and contexts.

Fig. 3. Virtual Mouse

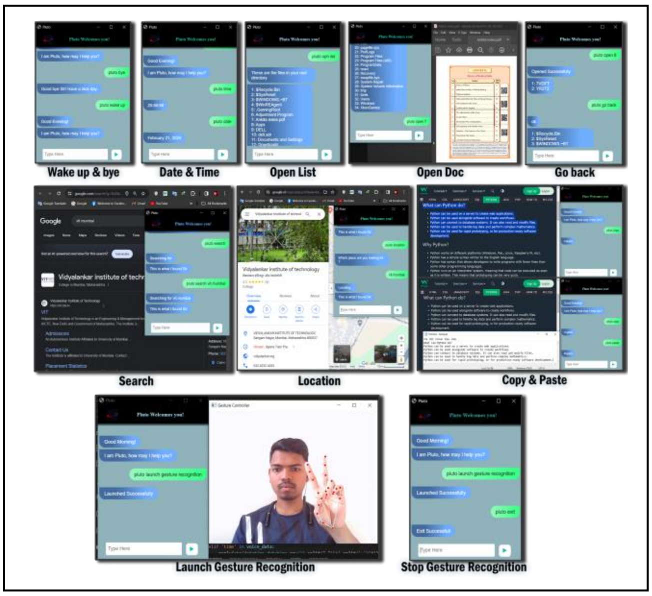


Fig. 4. Voice Assistance

Commands for voice assistances :

1) *Pluto Wake up*
2) *Pluto Launch Gesture Control*
3) *Pluto Search*
4) *Pluto Location*
5) *Pluto List*
6) *Pluto Open file*
7) *Pluto Copy*
8) *Pluto Paste*
9) *Pluto Date*
10) *Pluto Time*
11) *Pluto Bye*
12) *Pluto Exit*

V. Conclusion

The utilization of AI virtual mouse technology, which operates through hand gestures, presents an innovative and promising approach to computer interaction. By harnessing real-time camera input, we have devised a system capable of managing the mouse pointer without traditional input devices. This method offers users a more intuitive and accessible means of cursor control.

Moreover, when integrated with a voice assistant, the AI virtual mouse system can significantly enrich the user experience. Voice commands empower users to execute various tasks, including opening applications, navigating menus, conducting web searches, and controlling the cursor through hand gestures. As technology progresses, we anticipate the emergence of even more groundbreaking solutions that enhance user accessibility and experience.

Our voice assistant effectively fulfills user-specified tasks and boasts a wide range of functionalities, streamlining processes such as online searching. With the aim of potentially supplanting human server administrators, we have leveraged open-source software modules supported by the Anaconda community to construct this tool. Its modular design facilitates swift implementation of changes and integration of new features without disrupting system functionality.

VI. Future Scope

The proposed system faces certain limitations, including reduced precision in right-sided clicking operations of the mouse and challenges with smooth object dragging. Efforts are underway to address these limitations and enhance the system's functionality.

Furthermore, we are actively expanding our system by integrating voice assistant functionalities, which will contribute to a more intuitive Human-Computer Interaction (HCI) experience and propel its advancement..

Acknowledgments

We would like to express our sincere gratitude to our mentor, Prof. Rohit Barve for their guidance and support throughout this project. Prof. Rohit Barve was always available to answer our questions and provide feedback. We are grateful for their patience and encouragement.

We would also like to thank our classmates. For their collaboration and support. We worked together to overcome challenges and to achieve our goals. We shared resources and ideas, and we helped each other to stay on track. We are grateful for their friendship and support.

References

[1] Quam, D.L., et.al. (1990). Gesture Recognition with a Dataglove. In IEEE conference on Aerospace and Electronics (pp. 755-760).

[2] Guoli Wang., et.al. (2015). Optical Mouse Sensor-Based Laser Spot Tracking for HCI input, Proceedings of the Chinese Intelligent Systems Conference (pp. 329-340).

[3] Baldauf, M., and Frohlich, p. (2013). Supporting Hand Gesture Manipulation of Projected Content with mobile phones. In the European conference on computer vision (pp. 381-390).

[4] A. Haria, A. Subramanian, N. Asokkumar, S.Poddar, and J. S. Nayak. (2017). Hand gesture recognition for human computer interaction,‖ Procedia Computer Science, vol. 115, (pp. 367–374).

[5] Mayur, Yeshi., Pradeep, Kale., Bhushan, Yeshi., & Vinod Sonawane. (2016). Hand Gesture Recognition for Human-Computer Interaction. In the international journal of scientific development and research (pp. 9-13).

[6] Steven Raj, N., Veeresh Gobbur, S., Praveen., Rahul Patil., & Veerendra Naik. (2020). Implementing Hand Gesture Mouse Using OpenCV. In the International Research Journal of Engineering and Technology (pp. 4257-4261).

[7] Sneha, U., Monika, B., & Ashwini, M. (2013). Cursor Control System Using Hand Gesture Recognition. In the International Journal of Advanced Research in Computer and Communication Engineering (pp. 2278-1021).

[8] Varun, K.S., Puneeth, I., & Jacob, T.p. (2019). Virtual Mouse Implementation using OpenCV. In the International Conference on Trends in Electronics and Informatics (pp. 435-438)

[9] Shibly, K.H., Kumar, S., Islam, M.A., & Iftekhar Showrav, S. (2019). Design and Development of Hand Gesture Based Virtual Mouse. In the International Conference on Advances in Science, Engineering and Robotics Technology (pp. 1-5).I. Boglaev, “A numerical method for solving nonlinear integro-differential equations of Fredholm type,” *J. Comput. Math.*, vol. 34, no. 3, pp. 262–284, May 2016, doi: 10.4208/jcm.1512-m2015-0241.

[10] Abhik Banerjee, Abhirup Ghosh, and Koustuvmoni Bharadwaj. (2014)."Mouse Control Using a Web Came Based on Colour Detection", IJCTT, vol.9