Gesture Controlled Virtual Mouse With Voice Automation

Prithvi J (1BM19CS122)
S Shree Lakshmi(1BM19CS136)
Sohan R Kumar(1BM19CS159)
Suraj Nair(1BM19CS163)
Department of CSE

Ms. Sunayana S
Associate Professor,
Department of Computer Science and Engineering

Introduction

The field of Human-Computer Interaction has seen significant advancements with the introduction of innovative technologies. Traditional input methods such as keyboards, mice, and touchscreens have become more sophisticated, but still require direct contact with the computer, limiting the scope of interaction.

Gesture-based interaction has emerged as an alternative approach to traditional methods, and the Gesture Controlled Virtual Mouse is an innovative technology that enables intuitive interaction between humans and computers. This project presents a comprehensive study of the Gesture Controlled Virtual Mouse, which leverages state-of-the-art Machine Learning and Computer Vision algorithms to enable users to control input/output operations using hand gestures and voice commands without the need for direct contact.

The Gesture Controlled Virtual Mouse is designed using the latest technology and is capable of recognizing both static and dynamic hand gestures in addition to voice commands, making the interaction more natural and user-friendly.

The system does not require any additional hardware, and the implementation of the system is based on models such as the Convolutional Neural Network (CNN) implemented by MediaPipe running on top of pybind11.

The system comprises two modules, one of which operates directly on hands using MediaPipe hand detection, while the other module uses colored-caps of any uniform color. The system currently supports the Windows platform.

It is a mouse simulation system which performs all the functions performed by your mouse corresponding to your hand movements and gestures. Simply speaking, a camera captures your video and depending on your hand gestures, you can move the cursor and perform various mouse movements.

The project can be useful for various professional and non-professional presentations. It can also be used at home by users for recreational purposes like while watching movies or playing games.

Proposed System

The proposed Gesture Controlled Virtual Mouse system also includes a third module that leverages voice automation for wireless mouse assistance. This module allows users to perform mouse operations such as clicking, scrolling, and dragging, by simply giving voice commands. This feature is especially helpful for users who are unable to use hand gestures due to physical limitations. The voice automation module is implemented using state-of-the-art speech recognition algorithms that enable the system to accurately recognize the user's voice commands. The module is designed to work seamlessly with the other two modules of the system, allowing users to switch between hand gestures and voice commands effortlessly.

This module also adds a layer of convenience by allowing users to perform mouse operations from a distance, without the need for any direct contact with the computer. There are 2 Ways in which the user can control the mouse movements:

- Using Colored Caps
- Using Hand gestures

This makes it a useful tool for presentations, demonstrations, and other scenarios where the user needs to interact with the computer without being physically close to it.

Overall, the Gesture Controlled Virtual Mouse system is an innovative and user-friendly solution that simplifies human-computer interaction. With its advanced machine learning and computer vision algorithms, it offers a reliable and efficient way for users to control their computers using hand gestures, voice commands, or a combination of both.

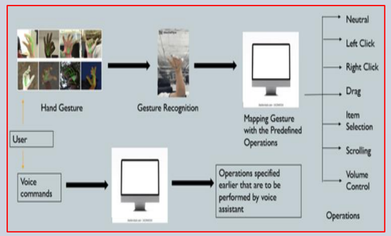


Fig 1: Proposed System(Gestured Controlled Mouse With Voice Assistant)

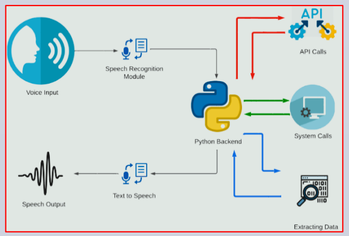


Fig 4:Overview of voice assistant(ECHO)

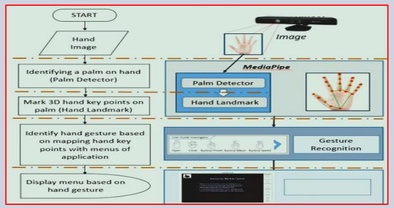


Fig 2: Flow Chart of virtual Mouse using Hand Gestures

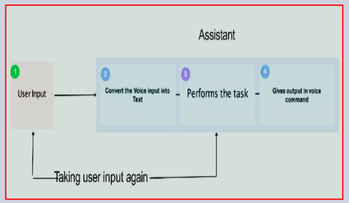


Fig 5: Flow Chart of Voice Assistant

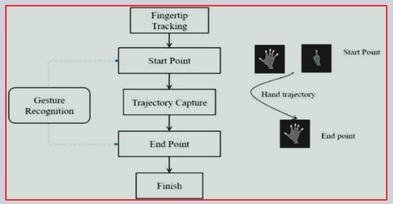


Fig 3:Flow Chart of Virtual Mouse using coloured caps

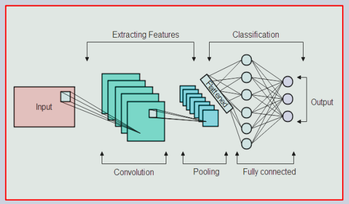


Fig 6:Classification Model Used-CNN

Methodology

Hand-Gesture Controlled Virtual Mouse

Operations of cursor-

- Hand Landmark detection(fingers up(1) and down(0))
- Move
- Left Click
- Double Click
- Right click
- Scroll Up
- Scroll down
- Volume Control
- Brightness Control

Steps-

1. It will detect the camera, video interface will be start
2. The camera can extract and recognize human hand gestures from video interface
3. Hand tracking functionality is done by mediaPipe
4. After the recognition the cursor move accordingly, to perform various operations

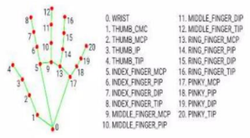
Classification Model

Deep Learning

- Deep Learning is a subset of machine learning. It is basically learning and improving on its own by examining other algorithm
- It works on artificial neural network that was designed to imitate human think and learn capabilities

Mediapipe

- MediaPipe to recognize the hand and the hand key points.
- MediaPipe returns a total of 21 key points for each detected hand.



Coloured Caps Controlled Virtual Mouse

Steps:

a. Obtain input video feed

b. Retrieve useful data from the image to be used as input

c. Filter the image and identify different colors.

d. Track the movement of colors in the video frame.

e. Implement it to the mouse interface of the computer according to predefined notions for mouse pointer control.

Voice Assistant(ECHO)

Speech Recognition module

The system uses Google's online speech recognition system for converting speech input to text. The speech input Users can obtain texts from the special corpora organized on the computer network server at the information centre from the microphone is temporarily stored in the system which is then sent to Google cloud for speech recognition. The equivalent text is then received and fed to the central processor.

Python Backend:

The python backend gets the output from the speech recognition module and then identifies whether the command or the speech output is an API Call and Context Extraction. The output is then sent back to the python backend to give the required output to the user.

API calls

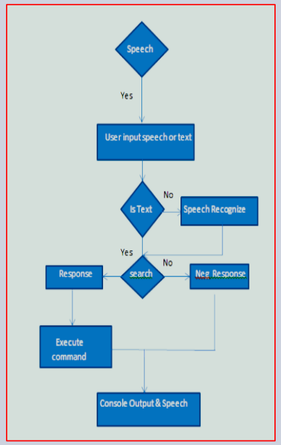
API stands for Application Programming Interface. An API is a software intermediary that allows two applications to talk to each other. In other words, an API is a messenger that delivers your request to the provider that you're requesting it from and then delivers the response back to you.

Content Extraction

Context extraction (CE) is the task of automatically extracting structured information from unstructured and/or semi-structured machine-readable documents. In most cases, this activity concerns processing human language texts using natural language processing (NLP). Recent activities in multimedia document processing like automatic annotation and content extraction out of images/audio/video could be seen as context extraction TEST RESULTS.

Text-to-speech module

Text-to-Speech (TTS) refers to the ability of computers to read text aloud. A TTS Engine converts written text to a phonemic representation, then converts the phonemic representation to waveforms that can be output as sound. TTS engines with different languages, dialects and specialized vocabularies are available through third-party publishers.



Results

The hand gesture and colored caps controlled virtual mouse along with voice recognition system incorporate various gestures like: neutral gesture, moving cursor, left click, right-click, double click, scrolling, drag and drop, multiple item selection. The voice assistant performs launch/stop gesture recognition, and content search on Google, identifies a location, navigates files, displays the current date and time, copies and pastes, sleeps/wakes up, and exit actions. The webcam is positioned at various distances from the user to monitor hand motions and gestures to detect fingertips, Gesture's ability is assessed under diverse lighting conditions such as bright light settings, low-light configurations, at a much farther distance from the camera, at a closer distance from the camera, with a left hand, right hand, both hands in camera, different backgrounds, and different hands of individuals of varying ages. The Voice Assistant is tested by providing diverse input via the mic and executing various functions such as location, file navigation, current time and date, copy and paste, sleep/wakeup, google search, and start and exit under various conditions.

Using hand gestures to control a mouse can increase productivity and ease of use, particularly for individuals with disabilities or those who find traditional mouse controls difficult. The automated training machine learning model accurately detects and classifies hand gestures, allowing for smooth and precise cursor control. While further research and testing may be necessary to optimize the system's performance, the results thus far suggest that a hand gesture-controlled mouse could become a valuable tool for computer users in the future.

Test Name	Cursor control
Test Description	To test cursor control functions.
Input	Camera image as input.
Expected Output	Cursor functions should be implemented using index and the middle finger.
Actual Output	Cursor functions are implemented accordingly.
Test Result	Success

Fig: Hand-Gesture Controlled Virtual Mouse Unit-Testing

Test Name	Cursor control
Test Description	To test cursor control functions.
Input	Camera image as input.
Expected Output	Cursor functions should be implemented using coloured Caps on finger tips.
Actual Output	Cursor functions are implemented accordingly.
Test Result	Success

Fig: Virtual Mouse Controlled Using Coloured Caps Unit-Testing

Test Name	control
Test Description	To Test Voice Assistant(ECHO's) Control Functions
Input	Voice/Text as Input
Expected Output	User requested functions should be implemented using ECHO
Actual Output	requested functions implemented accordingly
Test Result	Success

Fig: Voice Assistant ECHO Unit-Testing

For App's performance, following parameters are considered,
➢Start-Up: When the user clicks the first screen will be loaded in 5-10s seconds.
➢Usage with Other Apps: When the app is running in parallel with other apps, then there is no interference observed.
➢App's in background: When the app that is running in the background is retrieved, it has remained in the same state as it was before.
➢Surroundings: The Background does affect the app's performance

Fig: Application Performance

Output Snapshots

Fig: Using Hand-Gestures

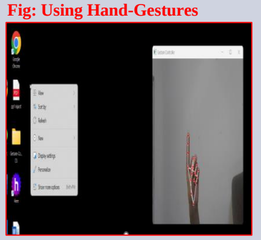


Fig: Voice Assistant ECHO

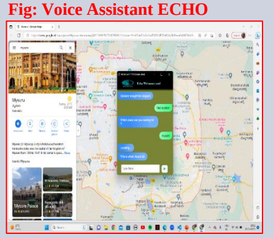


Fig: Using Colored Caps

Conclusion

In conclusion, Gesture Controlled Virtual Mouse is an innovative system that revolutionizes the way humans interact with computers. The use of hand gestures and voice commands provides a new level of convenience and ease to users, allowing them to control all I/O operations without any direct contact with the computer.

The system utilizes state-of-the-art Machine Learning and Computer Vision algorithms such as CNN implemented by MediaPipe running on top of pybind11 to recognize hand gestures and voice commands accurately and efficiently.

The two modules - one for direct hand detection and the other for colored caps of any uniform color to carter different user preferences and provide flexibility in usage. Additionally, the system incorporates a voice automation feature that serves various tasks with great efficiency, accuracy, and ease. With the current implementation of the system on the Windows platform.

Gesture Controlled Virtual Mouse presents an exciting prospect for the future of human-computer interaction. It is expected to increase productivity and convenience for users and could potentially have numerous practical applications in industries such as healthcare, gaming, and manufacturing.

The system has the potential of being a viable replacement for the computer mouse, however due to the constraints encountered; it cannot completely replace the computer mouse. The major constraint of the system is that it must be operated in a well-lit room. This is the main reason why the system cannot completely replace the computer mouse, since it is very common for computers to be used in outdoor environments with poor lighting condition.