International Journal of Aquatic Science
ISSN: 2008-8019
Vol 15, Issue 01, 2024

Empowering Huqman Computer Interaction Via Hand Gesture with Voice Assistant Integration

Pratiksha Dhakulkar[1], Vaishnavi Khadatkar[2], Pratiksha Kadu[3], Rohit Patil[4], Sahil Kakpure[5], Shreyas Balapure[6], Dr. S. W. Mohod[7]

[1,2,3,4,5,6]Students, Department of Computer Engineering, Bapurao Deshmukh College of Engineering, Sevagram, Wardha.
[7]Assist. Professor, Department of Computer Engineering, Bapurao Deshmukh College of Engineering, Sevagram, Wardha.

Email: [1]pratikshadhakulkar24@gmail.com, [2]khadatkarvaishnavi@gmail.com, [3]kadupratiksha92@gmail.com, [4]rohitpatil13092002@gmail.com, [5]sahilkakpure@gmail.com, [6]shreyashbalapure@gmail.com, [7]sudhirwamanrao@gmail.com

Abstract: The most popular invention in Human Computer Interaction (HCI) is the mouse. The most significant innovation in artificial intelligence that has the potential to drastically alter peoples' way of living is the voice assistant. Many devices are changing to become intelligent in their own distinct manners so they will interact with people easily. Thus, with the help of a hand, this research provides an innovative approach for controlling mouse movements using a real-time camera. With the help of system, human hand postures will be recognized more accurately in HCI applications, computing time will be reduced, and user comfort in regards to hand postures will improve. The application has good time-based performance based on the selected hand feature and recommended algorithm. The suggested positions for the hands along with the voice assistance help to make the system easier for the user. This system integrates image processing and deep learning concepts to allow users to operate their computers with simple hand gestures and spoken commands with help of OpenCV, MediaPipe, Python and other tools. Users may effortlessly switch between gestures and spoken commands for performing operations, access information, and interact with their devices when a voice assistant is incorporated.

Keyword: CNN (Convolutional Neural Network), Gesture Recognition, Mediapipe, Opencv, HCI (Human Computer Interface

1. INTRODUCTION

This system integrates image processing and deep learning concepts to allow users to operate their computers with simple hand gestures and spoken commands. Users may effortlessly switch between gestures and spoken commands for performing operations, access information, and interact with their devices when a voice assistant is incorporated.
Instead of a keyboard, mouse, or joystick, we can use hand gestures to control computer operations such as play/pause a media file, scrolling (left/right) scrolling (upward or downward) in a web page, and many more. We propose an emerging and efficient +computer interaction system based on a convolutional neural network (CNN) for effortlessly recognising

hand gesture movements. to recognise hand gestures and vocal instructions, the system employs computer vision and machine learning techniques that do not involve any additional hardware.
The voice assistant is an innovative concept that makes use of voice commands for interacting with digital devices and software applications. This innovative approach will contribute in changing the way we interact with technology, breaking down barriers while offering a more effortless and immersive user experience. It is also useful for speeding common tasks which include displaying the time and date, speeding Google searches, and detecting any location on a map, and other functions. Vocal command can be given to the voice assistants.
Usually, while making use of a wireless or Bluetooth mouse, a battery is required to power the mouse, whereas this research proposes a system, where the user uses their respective built-in camera or webcam and hand gestures to control the computer mouse operations. In the proposed approach, the web camera captures and then processes the captured frames, by acknowledging various hand gestures and hand tip movements prior executing the specific mouse operation. The Python programming language has been used in the development of the computer user interaction system, also OpenCV which is the library for computer vision is used in the computer user interaction system.

The proposed system aims to eliminate the use of traditional input devices such as mouse and key board to operate the laptop/PC. The goal is to implement additional gestures along with the two-hand gesture, so that user can perform more task efficiently. Instead of using more expensive sensors, a simple internal camera used to anidentify the gesture and perform the action. It helps the user to interact with a computer without any physical device to control the basis input operations The project overcomes the cursor operations in addition, the keyboard functions, play/pause, next/previous, zoom-in/zoom-out or other system operation and also perform some sort of operations with the voice command including web searches, locating files, etc.

2. LITERATURE SURVEY

M. Vasavi and Dr. D. N. Rao proposed a Gesture Recognition system. For User Interaction with Computer without Sensors using image processing, computer vision, deep learning techniques. ARM 11 processor and simple Haar-like features in viola and jones algorithm are used to make easier recognition and classification. High performance controller is developed using MINI 6410 board but it is restricted to arm 11.[1]
Dr. Jayant Nandwalkar and Mahima mandal proposed a system to control mouse using hand gesture and voice by RGB-D images and fingerprint detection. To recognise hand gestures and vocal commands, this research works with advanced machine learning and computer vision algorithms. It consists of two modules: one that works directly on the hand utilizing MediaPipe and another that works on the Windows platform using gloves of any uniform colour.[6]
Dr. Pratibha Waje designed Hand Gesture controller and voice Assistant using OpenCV, Machine Learning and Python. The controller had a potential to improve accessibility and convenience of computer interaction for users who prefer an alternative to traditional input devices. It was able to achieve 97% accuracy.[11]
Rohit Mukherjee and Pradeep Swethen proposed a system for Hand Gesture Controlled Laptop Using Arduino via mathematical algorithm. It was designed of three parts: an Arduino Uno, ultrasonic sensors, and a laptop. Ultrasonic sensors linked to the Arduino are used to determine gestures and the distance between the ultrasonic sensors and the hand. The main objective was

to reduce the effort required for communicating with computers via input devices by making use of simple gestures.[3]
S. Shriram and B. Nagaraj created a real-time AI virtual mouse system based on deep learning and computer vision, for the purpose of preventing the global propagation of COVID-19 It worked on the OpenCV library, which is the hand's computer vision library. Additionally, packages for Pynput, AutoPy, and PyAutoGUI has been used. It was able to reach an accuracy of 99% [4].
Kabid Hassan Shibly and Samrat Kumar Dey proposed a system to develop a Virtual Mouse using Gesture Recognition. The aim of this system is to control mouse cursor functions using only a simple camera instead of a traditional or regular mouse device. This system utilizes a color segmentation and detection technique to handle gestures that are captured using a webcam or built-in camera. The user can primarily use left, right, and double clicks in addition to various hand motions for up and down scrolling. This system is written in Python and makes use of the OpenCV package, which is based on computer vision.[2]
Prithvi J and S Shree Lakshmi proposed a gesture-controlled virtual mouse with voice using pybind11 that could correctly and effectively interpret hand motions and vocal commands. Additionally, the system has a speech automation capability that performs a variety of activities with remarkable ease, precision, and efficiency.[20]

3. PROPOSED METHOD

The library of OpenCV is used for computer vision, and the Media Pipe framework is used for hand tracking and gesture detection. The algorithm detects and identifies the hand movements and information with machine learning technique.
3.1 MediaPipe: Google provides an open-source tool called MediaPipe that may be applied in a machine learning pipeline. Cross-platform programming can benefit from the MediaPipe framework since time series data was used in its creation. Given its multimodal its nature, the MediaPipe framework is useful for a wide range of audio and video formats. Developers use the MediaPipe framework for developing and evaluating systems using graphs, and also to create devices for purposes related to applications. The pipeline configuration is used to execute out the stages involved in the MediaPipe-using system. Desktop and mobile scalability are made possible by the pipeline's compatibility across platforms.
The three primary components of the MediaPipe framework are performance evaluation, a framework for obtaining sensor data, and a set of reusable parts called calculators. A pipeline is a graph consisting up of units called calculators, each connected by streams through which data packets flow. Anywhere in the graph, developers may define or switch out custom calculators to develop applications of their own. Data-flow diagrams are produced by using streams and calculators. MediaPipe is used to create the graph, in which nodes are connected by streams and each of them is a calculator.
A hand or palm can be instantly detected and recognized with the single-shot detector model. The MediaPipe makes use of the single-shot detector concept. Due to the ease of learning palms, the hand detection module first learns for a palm detection model. In addition, the non-maximum suppression performs significantly better on small objects such fists or palms. As shown in Figure 1, a hand landmark model involves recognizing 21 joint or knuckle coordinates across the hand region.

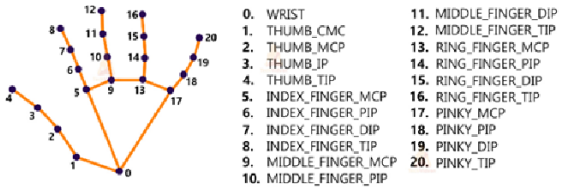


Fig. 1: Land Marks on the Hand

3.2 OpenCV: OpenCV is a computer vision library that comprises object recognition by image-processing algorithms, OpenCV is a Python programming language library that allows for the development of real-time computer vision applications. The OpenCV library is used in image and video processing, as well as analysis such as face and object detection.

The proposed system will eliminate the need for direct touch by employing Hand Gesture Recognition and a Voice Assistant. It consists of two main sections:
Hand Gesture Recognition:
We have manipulated click events and mouse cursor movement with the help of an internal camera. With the help of a voice assistant, nearly all basic I/O operations can be controlled with hand gestures, dynamic as well as static. The hands have been recorded in real time by a webcam.
Algorithm for Hand Gesture Recognition:
Step 1: Camera Initialization.
The computer user interaction system's camera. The frames that were collected by a laptop or PC's webcam act as the foundation for the proposed computer-user interaction system. The webcam will start recording video at once the video capture object is created via the Python computer vision library OpenCV. The frames are captured by the webcam and sent to the system.

Step2: Map Every Hand Gesture and Extract It.
Obtaining and analyzing the video. Each image is recorded by the webcam until the program ends, which is used by the computer-user interaction system. The process of locating hands in a video requires translating the images from BGR to RGB color space, as demonstrated by the code below:
Def findHands (self, img, draw= True):
Img RGB=cv2.cvtColor (img, cv2.COLOR_BGR2RGB)
self.results = self.hands.process (imgRGB)

Step 3: Matching the screen and camera resolutions.
Rectangular Region for Virtual Screen Matching—moving within the window. Using the transformational algorithm, the computer user interaction system moves the fingertip coordinates from the webcam screen to the full-screen computer window so that the mouse can be controlled. Once the hands are recognized and the finger assigned to that specific mouse function is identified, a rectangular box is drawn in relation to the computer window in the webcam region, allowing us to move the mouse cursor near the window.

Step 4: Identifying and executing the task based on the hand gesture.
Identifying Which Finger Is Up and Executing the Specific Mouse Function. At this point, we identify which finger is up with the help of the tip Id of the matching finger that we located

pwqwqeusing the MediaPipe and the appropriate coordinates of the up fingers. The mouse function is then executed according with this identification. Hand gestures and computer vision-based hand tip detection must be present for mouse function.

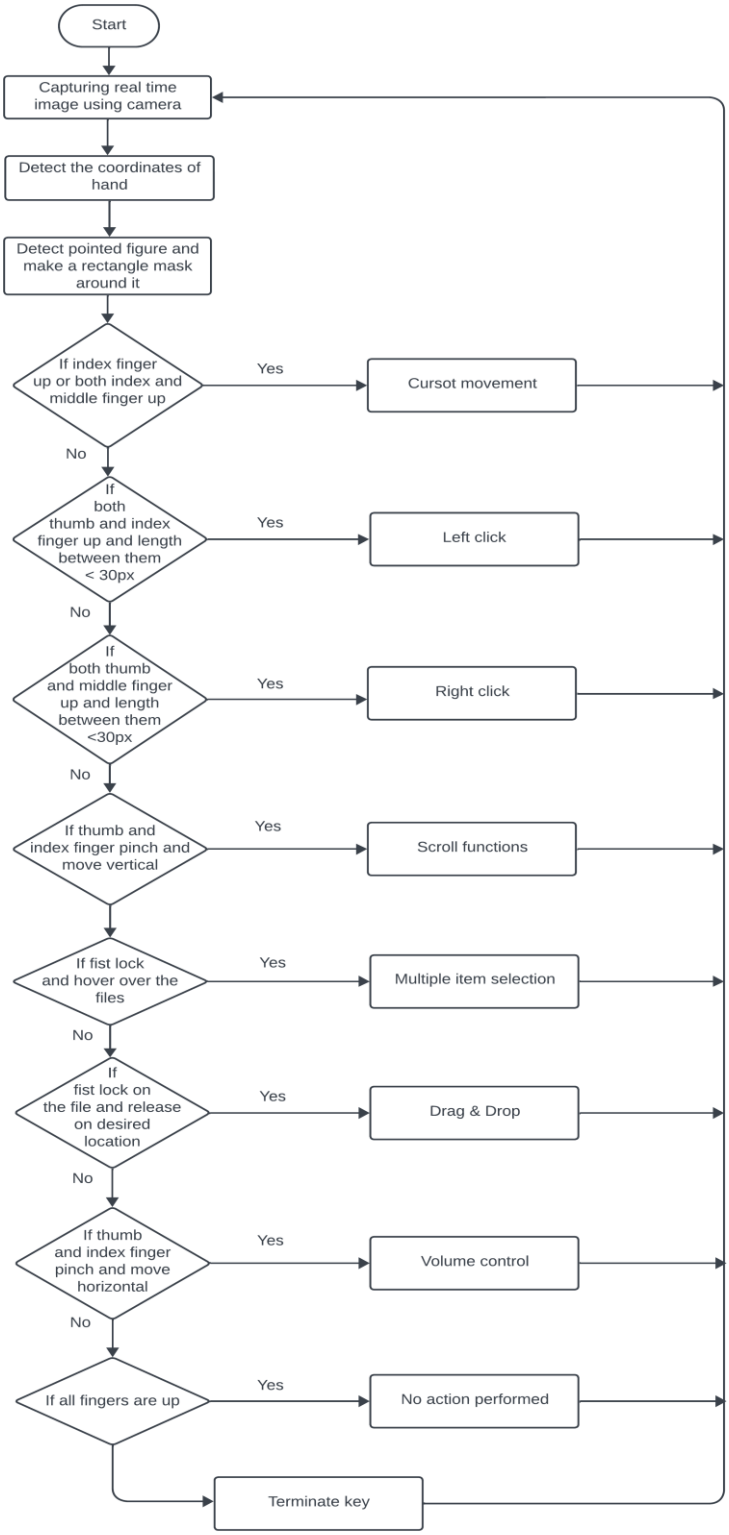
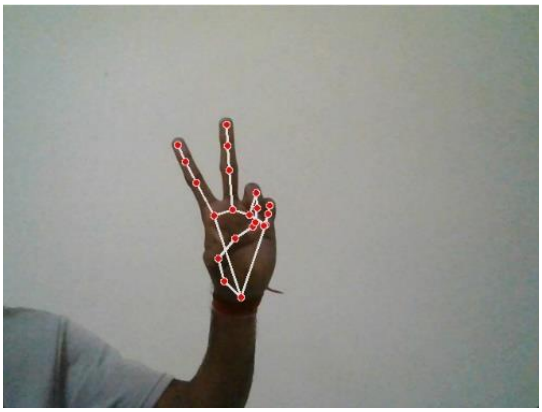


Fig. 2: Flowchart of Hand Gesture Functionalities

Following are the operations which performed by identifying the hand gestures for the particular operations using Computer Vision and Python Library (AutoPy).



Fig, 3: Gesture for Cursor Movement

Cursor movement: Cursor is operated by using index and middle finger tips. This gesture moves the cursor to desired location. Speed of the cursor movement is proportional to speed of hand.

Fig. 4: Gesture for Left Click

Left click: If both the thumb and index finger are up and length between them is less than 30 pixels, left click operation is performed.

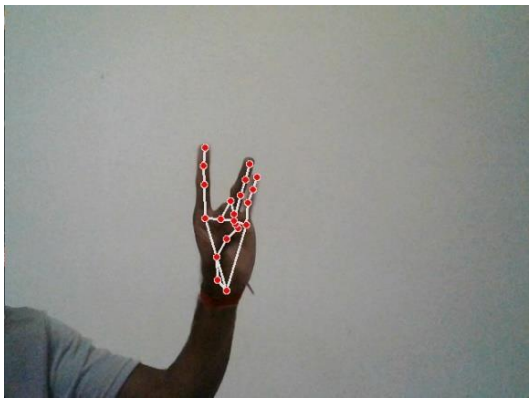


Fig. 5: Gesture for right click

Right click: If both the thumb and middle finger tips are up and length between them is less than 30 pixels, right click operation is performed.

Fig. 6: Fist Gesture for Multiple Item Selection

Drag & Drop: To perform drag and drop operation hold the file by making fist (drag) and release the fist on the desired location (drop).
Multiple item selection: To select multiple items lock a fist and hover over the files that has to be selected.

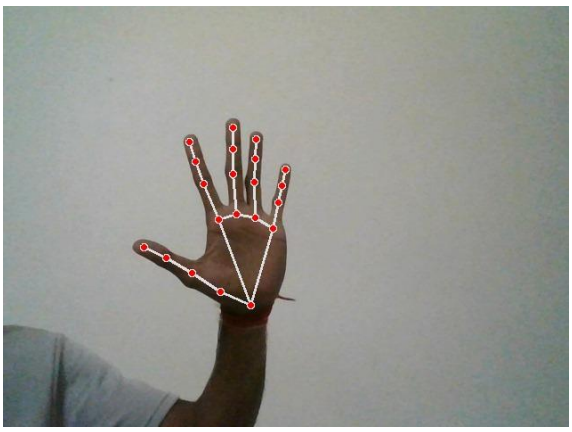


Fig. 7: Neutral Gesture

Neutral Gesture; Neutral gesture is used to halt/stop execution of current gesture by using all fingers up.
Scroll: If the tip of thumb and index finger form a pinch and moved in vertical direction (up/down) scroll function is performed.
Volume Control: If the tip of thumb and index finger forms a pinch and moved horizontally towards right and left, volume will increase and decrease respectively function is performed.

Voice Assistant:
A digital assistant that responds to specific voice commands and provides with appropriate data or performs functions at the user's request using recognition of speech, language processing algorithms, and voice synthesis.

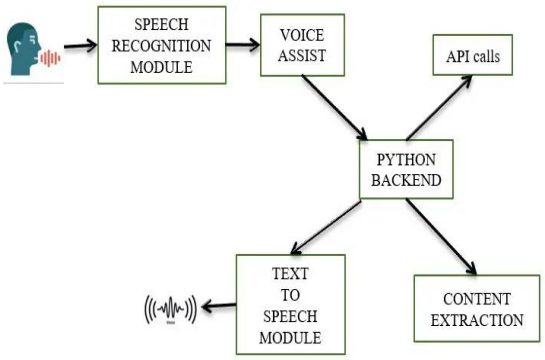


Fig. 8: Workflow for voice command

Algorithm for Voice Assistant:
Step 1: Recognition of Speech.
The system converts speech input to text via Google's online speech recognition system. The information centre's computer network server includes specific texts that users can access to obtain texts. The system saves audio from the microphone before transferring it to Google Cloud for recognition of speech.

Step 2: Python Backend
The speech or command output is parsed by the Python backend to identify if it is a System Call, Context Extraction, or API Call. In order to give the user, the intended outcomes, the output is subsequently delivered back to the Python backend. The acronym API for Application Programming Interface is used. A software interface known as an API facilitates communication between two apps. Stated differently, an API is the intermediary that transmits your request to the provider and then delivers the outcome.

Step 3: Content extraction
The process of extracting structured data from unstructured or semi-structured machine-readable materials is known as context extraction, or CE. This practice usually involves the use of natural language processing.to process text in human language.

Step 4: System Calls
A system calls is the procedure by which computer software requests the operating system's kernel for a service while it is executing. Examples of this include communication with essential kernel services like process scheduling, the creation and execution of new processes, and hardware-related services like accessing a hard drive. An interface between an operating system process and a process is provided via system calls.

Step 5: Text to speech
(TTS) is the ability of computers to read text aloud. Written text is transformed into a phonemic representation, which is then transformed into waveforms that a TTS Engine may produce as sound. TTS engines are available from third-party publishers in a range of languages, dialects and specialist vocabularies.

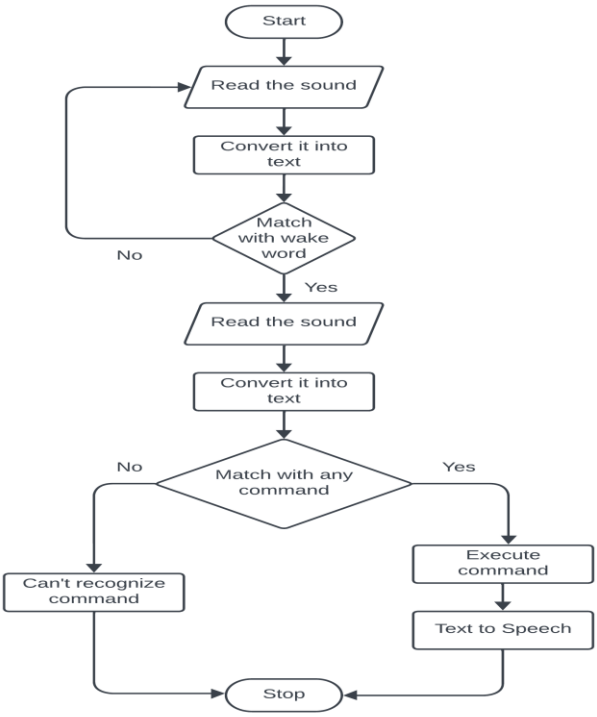


Fig. 9: Flowchart of Voice Assistant

This system is useful for many applications; it can be used to reduce the space for using the physical mouse, and it can be used in situations where we cannot use the physical mouse. system eliminates the usage of devices, and it improves the human-computer interaction.

Here are the three main commands that are working in the voice assistant.
1: Online searching commands
System takes commands to perform tasks such as play song, open opera, scan QR, send WhatsApp messages, open WhatsApp, google news, finding location, weather, open Gmail, open project report, search images, scan objects.
The commands for the same are as follows :
To play a song- “play song”
To open WhatsApp- “open WhatsApp”
To find location- “location”
To search images- “search image of”

2: Operating system commands
System takes commands to perform tasks such as adjusting the brightness level, open any application, shutting down a pc/laptop, opening a particular file, adjusting the volume, displaying date and time, changing the file name.
The commands for the same are as follows :
Changing Background- “change background”
Shutting Down- “shutdown system”
Adjusting Volume- “set volume level”
Open any Application- “open calculator”

3: Dynamic commands
We can use voice command to activate and deactivate the execution of hand gesture recognition Using “stop gesture recognition in query” command. This command will stop the execution of gesture recognition query.

Major applications:
1. This model has a far higher accuracy compared to other models that have been proposed as virtual mice.
2. Since touching the devices might result in a situation where the virus spreads via contact, the suggested method can be used for controlling mouse functions without a need for a physical mouse in the situations like COVID-19
3. Automation and robotic systems can be performed by the system without the need for additional devices.
4. Without a wireless or wired mouse, an AI virtual mouse can be utilized for playing games based on augmented reality and virtual reality.
5. The proposed system has functions in the field of robotics, particularly for controlling robots.
6. Utilizing the proposed system, it is possible to virtually design for prototyping purposes in architecture.

Table 1: Testing results

Hand Gesture	Function performs	Success	Failure	Approximation (%)
Both index finger and middle finger tip are up	Cursor Movement	100	0	100
Both index and index tip are up and distance <30px	Left click	99	1	99
Both index and middle tip are up and distance <30px	Right click	94	6	94
Both thumb and index are pinch are move vertically	Scroll function	99	1	99
If fist lock and hover over file	Multiple item selection	98	2	98
Fist lock and release	Drag & Drop	95	5	95
Both thumb and index tip pinch and move horizontally	Volume control	99	1	99
All fingers tip is up	No function performs	100	0	100
Result		784	16	98

4. RESULT AND DISCUSSION

A computer user interaction system that can be operated using hand gestures is an effective choice for people who have uncertainties using traditional mouse and keyboard. By simplifying computer interactions, this technology can improve their interaction with technological devices. Additionally, for people who prefer to work or play games without a physical mouse or touchpad, computer interaction systems which employ hand gestures can be useful. This

concept eliminates the requirement for a physical interface, allowing customers to operate their gadgets from a distance.

In various illumination condition and within different distances from the webcam for tracking, the hand gesture or landmark on the palm has been tested successfully. The summarization of results obtained from the experimental testing are shown in the Table 1. The testing has been performed under the various light condition by various persons to obtain the accuracy level of proposed system Compared to previous model, our model performed exceptionally well, with 98% accuracy.

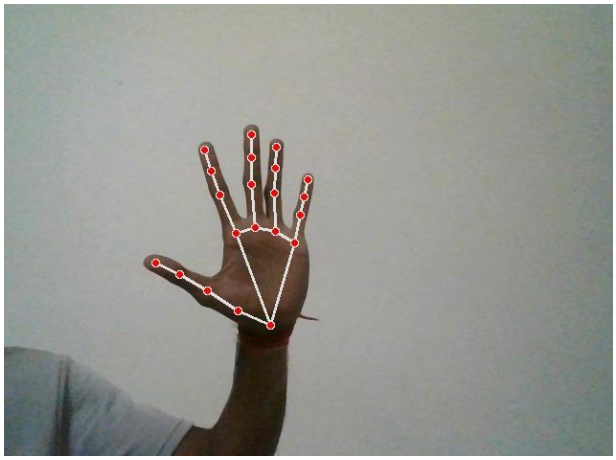


Fig. 10: Detection of hand landmarks

As comparison to current models of virtual mice, it is clear that a computer user interaction is overcome them all with respect to precision. The model that has proposed can essentially manage a laptop in a way that is comparable to a real mouse, executing the majority of mouse functions, such as Left/Right click events, drag and drop, volume control, multiple item selection, scrolling up and down functionalities. From Table 1, we see that the accuracy low Drag & Drop function as this is difficult to drop the file at particular due to at some point the connection between the system and hand terminated or the gesture by system not recognized and function not work properly at some instance. Also, accuracy is good and high for all the gesture and work properly.

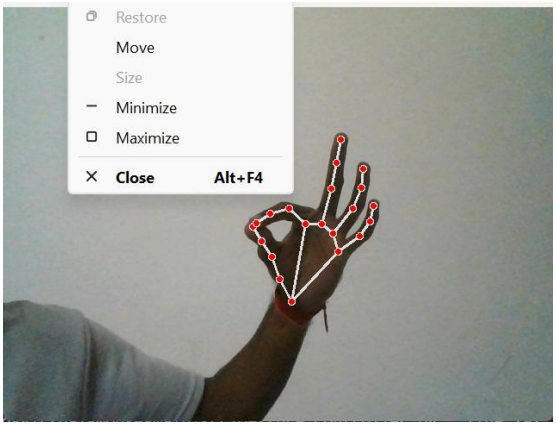


Fig. 11: Gesture to perform click event

Here are some of the examples and output of voice assistant, which can help you understand how the above processing works. It will do what is told. If the assistant can't understand what is told it will ask you to say that again please.

5. CONCLUSION AND FUTURE WORK

The compute user interaction system utilizes an internal camera to assist it locate and perform its task. We used the mouse for performing tasks like left, right, double click, and scrolling in along with selecting icons. This system utilizes motion detection and image comparison to move the mouse indicator and select the icon. Based on the analysis of the results, it can be predicted that the algorithms will work in any domain if we give them sufficient light and an effective camera. After that, our system will be more structured. Future plans call for the addition of extra functions like multi-window interaction, window expansion and reduction, window closing, etc. that can be managed with multiple fingers and the palm of the hand.
Hand gestures for computer user interaction may also enhance the user experience with more voice assistant support. Users will have even more control over their devices once the voice assistant is combined with the virtual mouse system. In addition to controlling the cursor with hand gestures on the screen, users can also give voice commands for a number of tasks, including opening apps, navigating menus, and performing web searches. As technology improves further, we will likely see even more creative solutions that boost accessibility and improve user experience.
Users can execute tasks hands-free and without the use of a keyboard or touchscreen with the aid of voice-based smart voice assistants. Additionally, they can speed up processes, increase accessibility, and improve the user experience in general. In order to understand voice instructions and provide precise and useful information in response, the project will use natural language processing (NLP) and machine learning methods.

This voice assistant currently works online and performs basic tasks like weather updates, streaming music, searching Wikipedia, playing music, opening desktop applications, etc. the system requires an internet connection. This Personal Assistant has been designed with ease of use as the main feature. The Assistant works properly to perform tasks given by the user. It will overcome the drawbacks of the existing solutions. A similar application is Cortana which comes with Windows operating system does not work in other operating systems but our system will work on any of the operating systems. There are many benefits of using this assistant such as it reduces labor costs, improves work quality, increases productivity, increases flexibility, etc.

Future work:
The computer user interaction system will provide more gestures in the future so that users can carry out numerous tasks with greater speed. The inadequacy in accuracy of the current model will overcome in future. As a result, future advancements to the current technique will make possible the use of both hands for diverse gestures. System will be able to serve more keyboard function in our future by making several advancements.

The Voice Assistant is currently implemented to be used on Windows, Android and iOS. Further this project can be implemented using any compact processing unit with dedicated Mic and Speaker. Further we can implement Home Automation like Controlling Light, AC, etc.

6. REFERENCES

[1] M. Vasavi PY. Ramesh Dr. D. N. Rao "Gesture Recognition for User Interaction with Computer Without Sensors" International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 [[2015]
[2] Kabid Hassan Shibly, Samrat Kumar Dey, Md. Aminul Islam, Shahriar Iftekhar Showrav "Design and Development of Hand Gesture Based Virtual Mouse" International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT) [2019]
[3] Rohit Mukherjee, Pradeep Swethen "Hand Gesture Controlled Laptop Using Arduino" International Journal of Management, Technology and Engineering [OCTOBER 2018]
[4] S. Shriram and B. Nagaraj "Deep Learning-Based real-time AI Virtual mouse system Using computer Vision to Avoid COVID-19 Spread" Hindawi Journal of Healthcare Engineering [2021]
[5] Pulkit Garg, Riya Shrivastava "VIRTUAL MOUSE USING HAND GESTURE" Journal of Emerging Technologies and Innovative Research (JETIR) [May 2019]
[6] Dr. Jayant Nandwalkar, Mahima Mandal "Control Mouse using Hand Gesture and Voice" International Journal for Research in Applied Science & Engineering Technology (IJRASET) [Apr 2023]
[7] Prof. Yuvraj V. Parkale, "GESTURE BASED OPERATING SYSTEM CONTROL" presented at Second International Conference on Advanced Computing & Communication Technologies [2012]
[8] Zepei Zheng, "Human Gesture Recognition in Computer Vision" Research SHS Web of Conferences 144, 03011 [2022]
[9] Jing Qi1,2 · Li Ma1,2 · Zhenchao Cui1,2 · Yushu Yu3 "Computer vision-based hand gesture recognition for human-robot interaction" Published on Springer [July 2023]
[10] B. Latha, Sri. Sowndarya "Hand Gesture and Voice Assistants" EDP Science [2023]
[11] Dr. Pratibha V. Waje, Ms. Shripanjali K. Gangurde, Ms. Snehal S. Sonwane "Hand Gesture (Virtual Mouse) and Voice Assistant Using OpenCV, ML, Python" International journal of Scientific research in Engineering and Management (IJSREM) [May 2023]
[12] Khushi Patel, Snehal Solanunde "Virtual mouse Using Hand Gestures and Voice Assistant" international Journal of Innovative Research in Technology [July 2022]
[13] Abhay Dekate, Chaitanya Kulkarni, Rohan Killedar Department of Computer Engineering, AISSMS College of Engineering, Pune, Maharashtra, India; "Study of Voice Controlled Personal Assistant Device" Article. December 2016
[14] Lokesh Gagnani, Hetvi Patel, Sakshi Chaturvedi, Rajeshwari Jaiswal Computer Engineering Department, LDRP Institute of Technology and Research Gandhinagar Gujarat India; "Gesture Controlled Mouse and Voice Assistant" Article. 10 Oct 2022
[15] Tran, D.-S., Ho, N.-H., Yang, H.-J., Kim, S.-H., & Lee, G. S. (2020). Real-time virtual mouse system using RGB-D images and fingertip detection. Multimedia Tools and Applications. doi:10.1007/s11042-020-10156-5
[16] 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
[17] Girish B G, Project Guide and Assistant Professor, Department of Computer Science and Engineering;" A Smart System using Hand Gestures and Voice" Article. June 2022
[18] D. L. Quam, "Gesture recognition with a Data Glove," IEEE Conference on Aerospace and Electronics, vol. 2, pp. 755–760, 1990. View at: Publisher Site | Google Scholar

[19] A. Banerjee, A. Ghosh, K. Bharadwaj, "Mouse Control using a Web Camera based on Color Detection", IJCTT, vol.9, March 2014
[20] Prithvi J, S Shree Lakshmi, Suraj Nair "Gesture controlled Virtual Mouse with Voice Automation" International Journal of Engineering Research and Technology (IJERT) [April 2023].
[21] Reddy, V. V., Dhyanchand, T., Krishna, G. V., & Maheshwaram, S. (2020). Virtual Mouse Control Using Colored Finger Tips and Hand Gesture Recognition.
[22] Barot, T., & Karhadkar, S. (2018). Development of Platform Independent Hand Gesture Controlled Wearable Mouse. 2018 3rd International Conference for
[23] Convergence in Technology (I2CT).