International Research Journal of Multidisciplinary Scope (IRJMS), 2024; 5(1):26-35 IRJMS

Original Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2024.v05i01.0127

Gesture Controlled Virtual Mouse with Voice Assistant

Ismail Khan, Vidhyut Kanchan, Sakshi Bharambe*, Ayush Thada, Rohini Patil

Terna Engineering College, Nerul, Navi Mumbai, India *Corresponding Author's Email: sakshibharambe3@gmail.com

Abstract

The usage of hand gestures to operate digital devices has grown in recent years, and developing accurate gesture recognition systems using computer vision technology is becoming easier. Voice assistants have also become more common, with developments in artificial intelligence (AI) and natural language processing (NLP) allowing for more accurate voice recognition. Our goal is to create a system which combines both gesture-controlled mouse functionality and voice assistant to execute things that a regular mouse and keyboard can. The GCVA (Gesture Controlled Voice Assistant) comes with a gesture-controlled mouse and a voice assistant. A gesture-controlled mouse uses OpenCV and Media-Pipe, whereas a voice assistant uses AI and NLP concepts to recognize the voice commands given by the user. The system should be used by performing hand gestures and executing the right click, left click, drag, drop, volume control and computer cursor functions, thereby omitting the need for a physical mouse. The success of this system is observed in terms of practicality, stability, and compatibility with physical mechanisms. The voice assistant works and responds to pre-defined commands present in the system. The GCVA system was tested by four different users in order to generate the accuracy scores of mouse gestures and response time graphs for voice assistant. The system yields an accuracy rate of 98%. The system's accuracy and usability showed that hand gesture detection and voice assistants have the potential to replace traditional input devices in the future.

Keywords: Human Computer Interaction, Computer vision, Artificial Intelligence, Voice Assistant, Media-Pipe.

Introduction

With the advancement in technology, the devices have become wireless and some of them do not exist. Users can perform the functions of those physical devices by using their personalized gestures. Hand gestures could be used as a communication tool by humans. By utilizing a laptop or computer with a web-camera and microphone, the system can perform the operations without the need for any equipment to control it. These gestures allow humans to communicate in the environment via hand shaking, thumbs up and down, and hand signs. According to an analysis of the human environment, hand gestures are one of the most useful and natural communication methods.

A Virtual mouse provides the benefit of controlling the system without the use of a physical mouse. Users can use their hand to operate the system with some pre-defined hand gestures, where each gesture will perform a certain action and the user can perform tasks with their hand instead of using a traditional mouse. A voice assistant helps the user to operate the system using their voice to execute pre-existing commands in the system.

Using AI with Python and OpenCV to create a Gesture Controlled Virtual Mouse and voice assistant system makes it easy for users to operate the system with their hands using different gestures and an integrated voice assistant system that will let users control the system using their voice by providing various voice commands (1). In an era where artificial intelligence (AI) is progressively indispensable for tackling intricate challenges in sectors ranging from healthcare and entertainment to finance and education, AI emerges as a testament to its versatility. The advent of the PC mouse has been a pivotal development in the field of Human-Computer Interaction (HCI), revolutionizing the way individuals engage with technology (2). The human hand is capable of transmitting large amounts of information via typing or sign language. It can form a great number of complex poses and is different from person to person in terms of colour, size, and shape (3).

(Received 13th October 2023; Accepted 02nd January 2024; Published 30th January 2024)

Related work

The system produces mouse movement by utilizing certain packages, such as 'mouse'. A virtual mouse system based on HCI cursor movements (4). Various functions were controlled by the hands with coloured caps on fingertips. Varun *et al.,* (5), developed an application for preschoolers by using a comparison of the two libraries. Shibly *et al.,* (6), designed a system that uses visual landmarks to identify the regions of interest (ROI) and cross-references them with those regions using computerized algorithms.

A speech synthesizer that generates the spoken language based on the written input (7). GDPR (General Data Protection Regulation) system (8), which helps to protect personal information and speech data. Haria *et al.,* (9), designed a machine learning based system that compares different classification algorithms such as K-nearest neighbour (K-NN), decision tree, support vector machine (SVM), and artificial neural networks (ANN) for recognizing hand gestures and evaluates the performance of the system using various metrics such as accuracy, precision, recall, and F1-score. Various approaches for recognition of hand gestures, including the use of color, depth, and skeleton-based models (10). A system built with the Boof CV library, which has an object tracking-based virtual mouse application to track the movements of hand (11). A voice assistant system that will collect the audio from the microphone and transform it into text before sending it through GTTS (Google text to speech) (12). The GTTS engine will transform text into an audio file in English, which will then be played using the 'play sound package'. Shriram *et al.,* (13), proposed a system that recognizes gestures using a convolutional neural network (CNN). The paper evaluates the system's performance and reports a gesture recognition accuracy rate of 96.2%. Khan *et al.,* (14), proposed a system that captures and processes the video from a webcam and then isolates the user's hand using a background subtraction method. The system also implements marker motion tracking to determine the location and movement of the hand. Rajendra *et al.,* (15), developed software for mouse movements. The author used a used green coloured object for performing cursor movement with the mouse.

Methodology

The proposed GCVA system includes a gesture-controlled virtual mouse and a voice assistant, both of which operate simultaneously. OpenCV, Media Pipe, NLP, and Python modules were used to build the system.

The below steps demonstrate the workings of the GCVA system.

Step 1: The user will provide input via gesture or speech.

Step 2: If the input is in the form of a gesture, the gesture recognition mechanism will be activated.

Step 3: Using OpenCV and Media Pipe, the function will map the coordinates on the hand, referred to as landmarks.

Each gesture has a distinct hand landmark; these landmarks are used to detect the position of the hand.

Step 4: Based on the gesture detected, the system executes the desired function.

Step 5: When a voice command is provided, the system checks whether it is a command for gesture or not; if yes, then it launches the gesture recognition mechanism and repeats steps 3 and 4.

Step 6: Aside from the gesture command, the voice assistant will analyse the commands given by the user as an input and respond accordingly.

The below Fig 1 represents the GCVA system and its working.

Gesture Controlled Voice Assistant (GCVA) system is divided into two phases.

Phase 1: Gesture Recognition

For gesture recognition, a webcam is used to capture data for the hand, and the 'multi_hand_landmarks' library is used for mapping the 21 landmarks on the hand. This data is used to find the position of hands, the distance between the fingers, and to check whether the fingers are up or down. Based on the gesture detected, the system executes the desired function. The below Table 1 represents the methods used for gesture recognition and their roles.

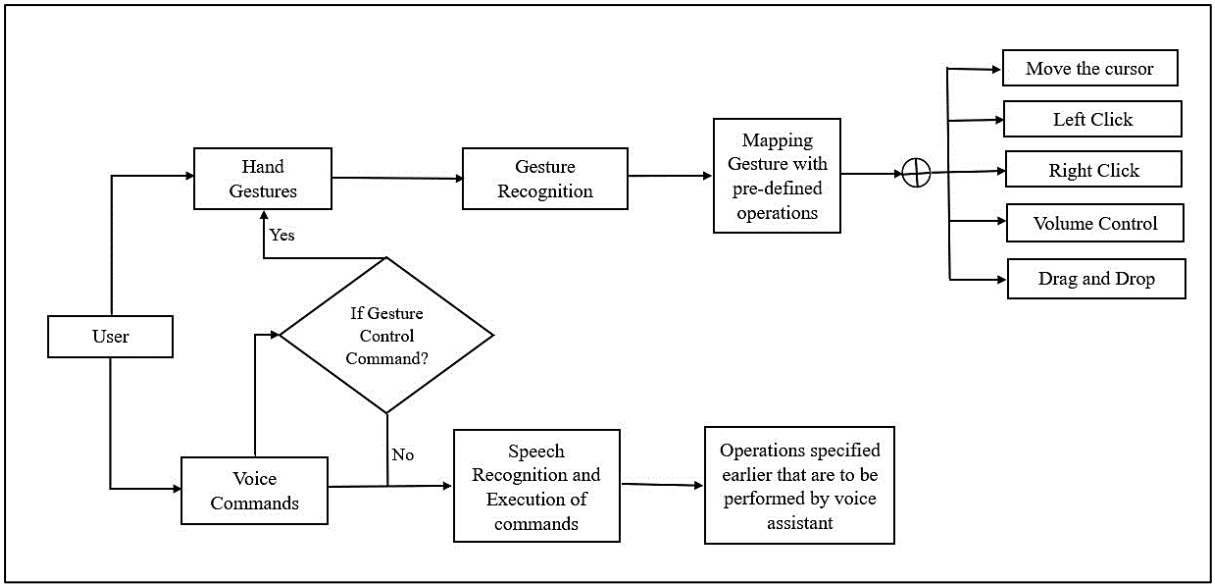


Figure 1: GCVA System

Table 1: Methods used for gesture recognition and their roles

Methods used	Role
mediapipe.solutions.hands	It is used for detecting hands from the webcam
multi_hand_landmarks	This function is used for mapping the 21 landmarks on the hand
mediapipe.solutions.drawing_utils	It is used to draw connections between landmarks over the detectedhand
findPosition	This function is used to find theposition of the hand in the window
findHands	This function is used to recognize thehand gestures
fingersUp	This function is used to checkwhether the fingers are up or down
findDistance	It is used to find the distance betweenthe fingers

Table 2: Methods used for voice assistant and their roles

Methods used	Role
recognizer	The purpose of a Recognizer function is to recognize speech
pyttsx3.init	It is used to convert the entered text into speech
record_audio	This function records user command
respond	This function gives response back to the user alongwith the desired action

The libraries used by Gesture Recognition are:

1. OpenCV: OpenCV is a free and open-source computer vision library that uses NumPy to perform high-level mathematical operations on multidimensional arrays and matrices.

2. Media-Pipe: Google Media-Pipe is an open-source framework for creating real-time computer vision, machine learning, and audio processing apps.

Phase 2: Voice Assistant

The AI implemented voice assistant will take the user's speech as input and based on the command received, it will analyze and execute the desired function. This could also launch the gesture recognition function. Table 2 represents the methods used for voice assistant and their roles.

The libraries used by voice assistant are:

1. **PyAutoGUI:** PyAutoGUI is a Python package that can be used to manage mouse cursor movements, clicks, and keyboard presses. It can be used for GUI testing, automation, and game creation.
2. **Pyttsx3:** Pyttsx3 is a Python package that converts text to speech using Text-To-Speech engines. It is installable with pip and can be used for assistive technologies, language learning, and speech-enabled applications.

The GCVA System is an integrated system divided into two phases. It determines if the voice assistant is active and performs actions based on user commands. If the command is to activate gesture recognition, it activates the function and performs the functions based on the gesture. The Fig 2 and Fig 3 provides a detailed explanation of the system's operation and functions.

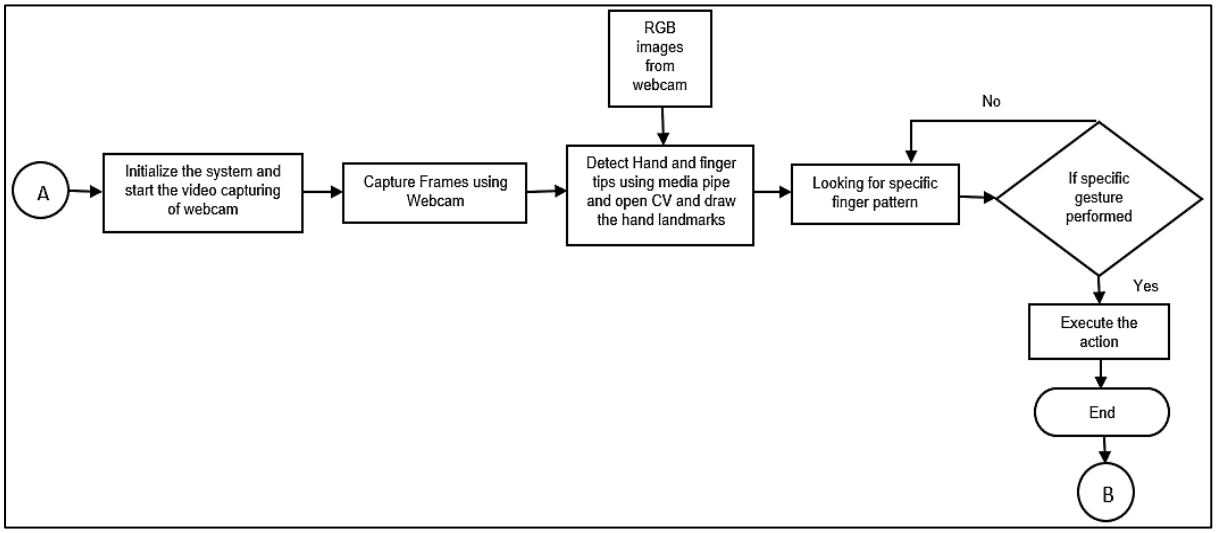


Figure 2: The Detailed Flowchart of the GCVA system – I

Result and Discussion

By using the libraries, the following gestures are performed in GCVA using various functions, which are discussed in the below Table 3 along with their actions.

After using these gestures on GCVA, the following results are generated: Fig 4 represents the mapping of the coordinates on the palm; Fig 5 shows the left click; Fig 6 represents the right click operation; Fig 7 shows the GUI of the voice assistant and Fig 8 shows the working of voice assistant.

Performance Analysis

GCVA PHASE I

The Accuracy score of every gesture performed

To analyze the result shown in Table 5, an experimental test was performed. The test was carried out 25 times by four people for seven different functions, resulting in 700 gestures with manual labelling. The tested gestures are: Mouse movement, Left Click, Right Click, Volume Control, Drag and drop, Double Click and exit function. Table 4 below shows the accuracy of different gestures performed. The Fig 9 shows the accuracy graph of different Gestures performed.

The Table 4 shows the accuracy of the GCVA system for different gestures performed, where the mouse movements, volume control, double click and exit functions show a maximum

accuracy of 100%, while the left click and right click functions show an accuracy of 98% and 96%, respectively. The minimum accuracy of 91% was shown by the drag and drop function.

GCVA PHASE II

To test the response time of the system, 4 different users were included in our system. The 7 questions asked to the assistant by 4 different users are related to time, date, google my location, google search on any website, launching gesture recognition, stopping gesture recognition, and exiting voice assistant. The tested results are measured in seconds. Table 5 shows the response time of 4 different users for 7 different questions asked by them to the assistant. The Fig 10 shows the corresponding response time graph.

The Table and figure show the response time of 7 different questions tested with 4 different users. User 1 and user 3 get the minimum response time of 1.91sec. for Que 1. The minimum response time of 4.51sec. and 44.25 sec was observed by user 3 for Que 2 and Que 7, respectively. For Que 3, the minimum response time observed by user 1 was 10.88 sec. For Que 4, 5 and 6, the user 2 observed a minimum response time of 18.55 sec, 27.67 sec and 36.73 sec respectively.

Accuracy Validation with contemporary researcher

The accuracy of different models is compared with the accuracy of the GCVA system in Table 6 and the comparison graph is formed with the tabular data in Fig 11. The table shows that the accuracy of the GCVA system is 98% compared to different models, where K. H. Shibly and A. Haria's system has an accuracy score of 78% and Tran, DS.'s system has an accuracy score of 93.25%.

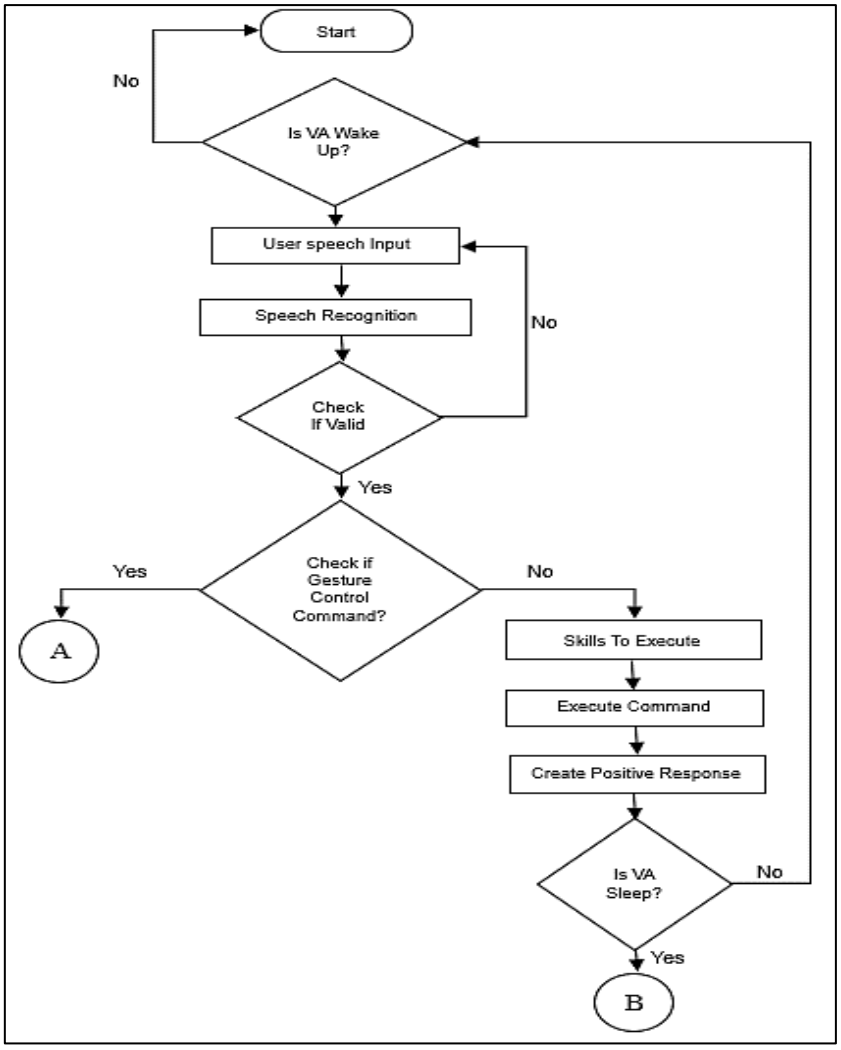
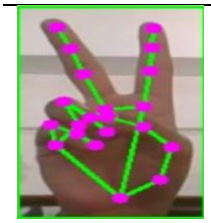
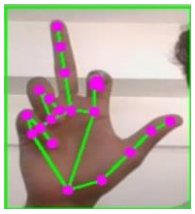
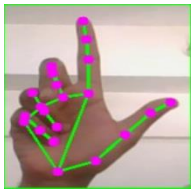
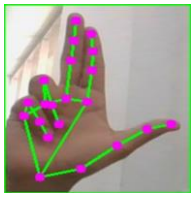
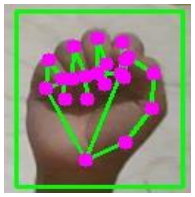
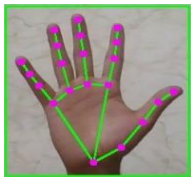
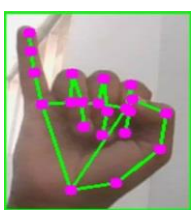


Figure 3: The Detailed Flowchart of the GCVA system – II

Table 3: Gestures and their actions

Gestures	Actions
	**Moving Cursor:** This gesture is used for the movement of the cursor on the Screen.
	**Left Click:** This gesture is used to perform the left click operation.
	**Right Click:** This gesture is used to perform the right click operation.
	**Double Click:** This gesture is used to perform the double click operation.
	**Drag:** This gesture is used to select and move the files across the screen.
	**Drop:** This gesture is used to release the selected file at the required location.
	**Volume Control:** This gesture is used to control the volume of the system.
	**Exit Function:** This gesture is used to exit the Gesture Recognition Function

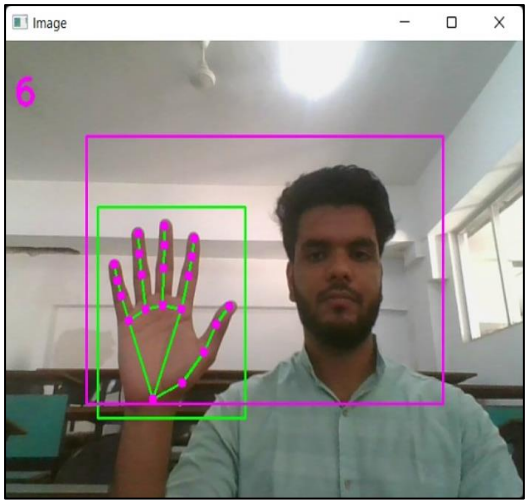


Figure 4: Mapping of coordinates on Palm

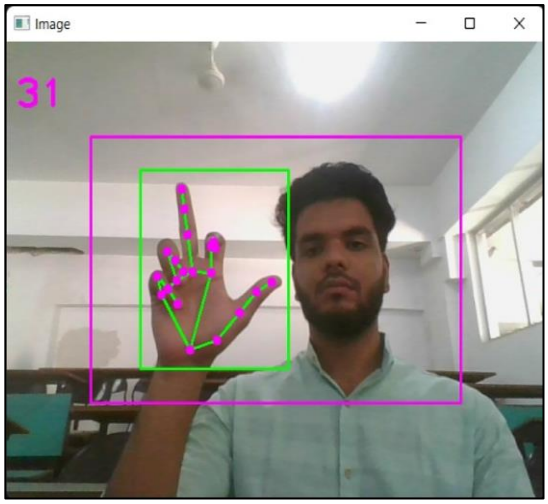


Figure 5: Left Click

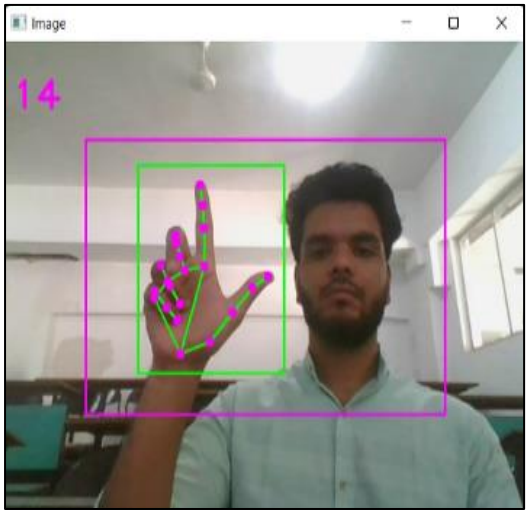


Figure 6: Right Click

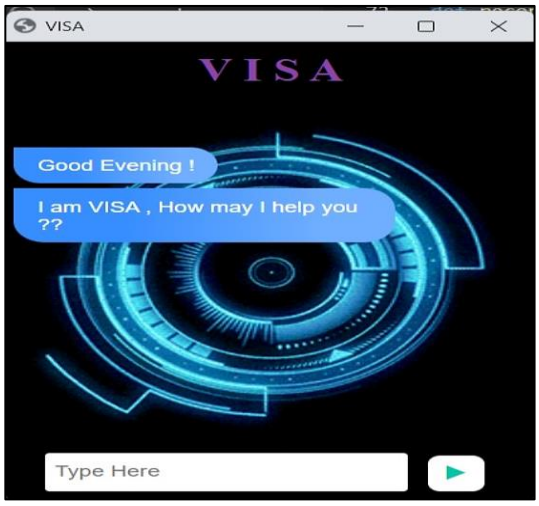


Figure 7: Voice assistant GUI

Figure 8: Working of Voice Assistant

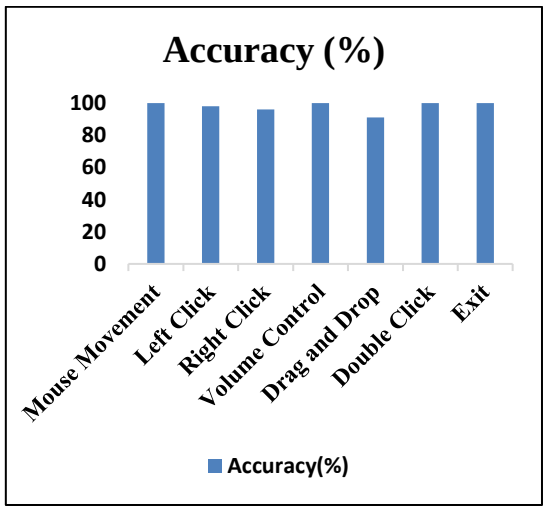


Figure 9: Accuracy graph of different Gestures performed

Table 4: Gestures and their accuracies

Mouse Functions Performed	Accuracy (%)
Mouse Movement	100
Left Click	98
Right Click	96
Volume Control	100
Drag and Drop	91
Double Click	100
Exit	100

Table 6: Comparison of accuracy with accuracy of different papers

Existing models	Accuracy (%)
Ref (10)	93.25
Ref (9)	78
Ref (6)	78
GCVA System	98

Table 5: Response time of 4 different users

Users	Response Time (in sec)						
	Que 1	Que 2	Que 3	Que 4	Que 5	Que 6	Que 7
User-1	1.91	4.53	10.88	22.44	31.59	39.9	46.38
User-2	2.21	4.94	11.33	18.55	27.67	36.73	47.32
User-3	1.91	4.51	13.08	19.6	29.05	38.18	44.25
User-4	2.02	4.68	11.38	27.52	36.75	47.23	53.62

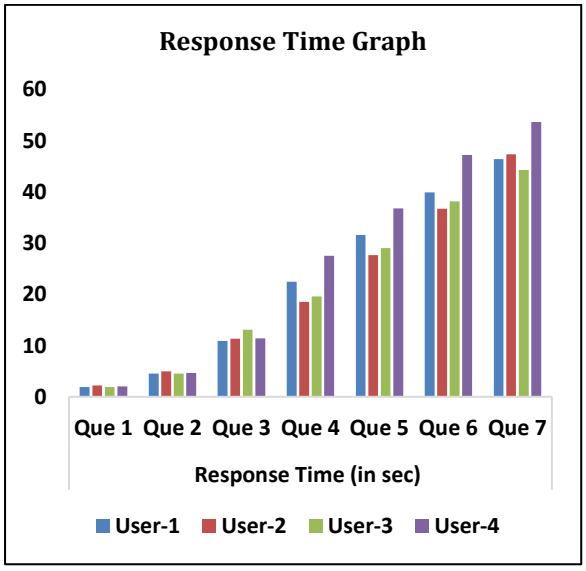


Figure 10: Response Time (in sec)

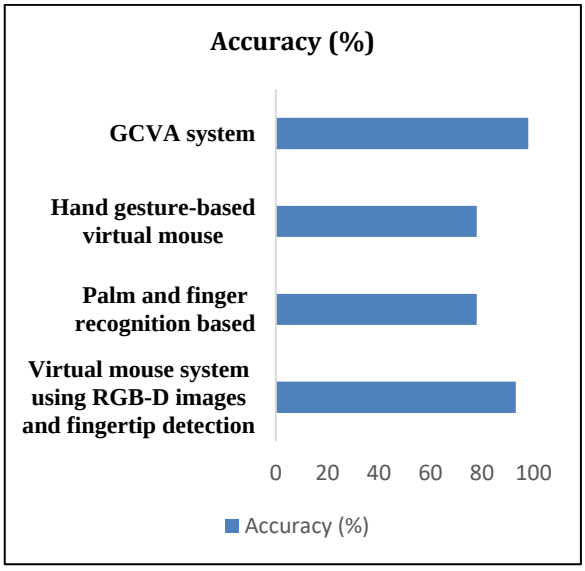


Figure 11: Comparison graph of different models

Conclusion

The GCVA system was implemented using Python and its libraries. This will let the user operate the system without the use of a hardware mouse just by using hand gestures along with a voice assistant, which could perform the actions of a physical mouse and dynamic functions like telling time, date, and web search, respectively. This will make the system completely hands-free and reduce the need for and cost of a physical mouse. In comparison to other systems, the GCVA system performs better, according to the observations.

Abbreviations

Artificial Intelligence (AI), Natural Language Processing (NLP), Gesture Controlled Voice Assistant (GCVA), Open Source Computer Vision (OpenCV), Human-Computer Interaction (HCI), Regions of Interest (ROI), General Data Protection Regulation (GDPR), K-nearest neighbour (K-NN), Support Vector Machine (SVM), Artificial Neural Networks (ANN), Computer Vision (CV), Google text to speech (GTTS), Convolutional Neural Network (CNN), Numerical Python (NumPy), Graphical user interface (GUI), Seconds (sec), Question (Que) and Reference (Ref).

Acknowledgement

The authors wish to acknowledge help of Dr. Anant Patil in revision of the manuscript based on the reviewers comments.

Author's contribution

IK: Conceptualization and Implementation, Methodology Performance evaluation; VK: Implementation, Formal analysis, performed the simulation work and results generated, Review and Editing; SB: Methodology , Writing Draft ,reviewing the literature, Review and Editing, revision; AT: Conceptualization, Methodology, Writing Draft, paper formatting , Review and Editing; RP: Conceptualization, Methodology, Revised Manuscript Review, Editing and approval; All the authors read and approved the final version of the manuscript.

Conflict of interest

None

Funding for the study

None

Ethical approval

Not applicable

Reference

1. Kavitha M, Umesh KV, Velaga K, Jallipalli JANK and Yellamma P. editors. Simulated Mouse Mechanism using Hand Movement Signals. Proceedings of the International Conference on Applied Artificial Intelligence and Computing (ICAAIC 2022), 09-11 May 2022, Salem. India: IEEE; 2022.

2. Guha J, Kumari S, Verma SK. AI Virtual Mouse Using Hand Gesture Recognition. International Journal for Research in Applied Science & Engineering Technology. 2022;10(3):3070-76.

3. Križnar V, Leskovšek M, Batagelj B. editors. Use of Computer Vision Based Hand Tracking in Educational Environments: 44th International Convention on Information, Communication and Electronic Technology (MIPRO), 27 September 2021 - 01 October 2021, Opatija. Croatia: IEEE; 2021.

4. Neogi D, Das N, Deb S. editors. BLINK-CON: A HANDS FREE MOUSE POINTER CONTROL WITH EYE GAZE TRACKING: Mysore Sub Section International Conference (MysuruCon), 24-25 October 2021, Hassan. India: IEEE; 2021.

5. Varun KS, Puneeth I, Jacob TP. editors. Virtual Mouse Implementation using Open CV: Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019), 23-25 April 2019, Tirunelveli. India: IEEE; 2019.

6. Shibly KH, Kumar Dey S, Islam MA, Iftekhar Showrav S. editors. Design and Development of Hand Gesture Based Virtual Mouse: 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT 2019), 03-05 May 2019, Dhaka. Bangladesh: IEEE; 2019.

7. KN, RV, SSS, DR. editors. Intelligent Personal Assistant - Implementing Voice Commands enabling Speech Recognition: International Conference on System, Computation, Automation and Networking (ICSCAN 2020), 03-04 July 2020, Pondicherry. India: IEEE; 2020.

8. Christensen AT, Olesen H, Sørensen L. editors. Digital Voice Assistants: A new kind of user agent: 13th CMI Conference on Cybersecurity and Privacy (CMI) - Digital Transformation - Potentials and Challenges(51275), 26-27 November 2020, Copenhagen. Denmark: IEEE; 2020.

9. Haria A, Subramanian A, Asokkumar N, Poddar S, Nayak JS. editors. Hand gesture recognition for human computer interaction: 7th International Conference on Advances in Computing & Communications (ICACC 2017), 22-24 August 2017, Cochin. India: ELSEVIER; 2017.

10. Tran DS., Ho NH, Yang, H. Real-time virtual mouse system using RGB-D images and fingertip detection. Multimedia Tools and Applications.2021; 80:10473–490.

11. Roy N. Controlling Mouse Cursor Based on Image Processing and Object Detection. International Journal of Engineering Science and Computing.2017;7(8):14762-63.

12. Subhash S, Ullas A, Srivatsa N, Santhosh B, Siddesh S. editors. Artificial Intelligence-Based Voice Assistant: 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), 27-28 July 2020, London. UK: IEEE; 2020.

13. Shriram S, Nagaraj B, Jaya J, Shankar S, Ajay P. Deep Learning-Based Real-Time AI Virtual Mouse System Using Computer Vision to Avoid COVID-19 Spread. Journal of Healthcare Engineering,2021; 2021.

14. Khan F, Halim B, Rahman A. Computer Vision Based Mouse Control Using Object Detection and Marker Motion Tracking. International Journal of Computer Science and Mobile Computing. 2020; 9(5):35-45.

15. Rajendra N, Aishwarya R, Roopa Shree NK, Vibha C, Rao B. Cursor Movement by Object Detection Based on Image Processing. International Journal of Research Publication and Reviews. 2021;2(5):18-22.