IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 14, May 2023

# Gesture Virtual Mouse and Voice Assistant using Machine Learning

**Prof. Saranya Raj[1], Aman Nautiyal[2], Pratham Sharma[3], Nishant Sharma[4]**
[1,2,3,4]Computer Science and Engineering
Dronacharya Group of Institutions, Greater Noida, India.

**Abstract:** *By controlling cursor movement with a real-time camera and microphone, this project advances the Human Computer Interaction (HCI) paradigm in the field of computer science. The hand movement and speech is the most effortless and primitive way of communication. It's a replacement for the present ways, which entail manually moving a physical computer mouse or pressing buttons. Instead, the system controls and performs numerous mouse activities using a camera for computer vision technology and a microphone for speech recognition and processing. It can perform all functions that a physical mouse can. The Virtual Mouse continuously gathers real-time visuals and voice commands, which are then filtered and converted in a number of steps. When the procedure is completed, the programme uses image processing and natural language processing to extract the valid command needed to complete the task. Specially abled people with hand problems can use this virtual mouse to control the computer's mouse functionalities.*

**Keywords:** *Virtual Mouse , Voice Assistant , OpenCV , MediaPipe , Python*

## I. INTRODUCTION

The project's main goal is to create a hands-free virtual Mouse system that focuses on a few key applications in development. This project aims to eliminate the need for a physical mouse while allowing users to interact with the computer system via webcam and speech using various image and audio processing techniques. This project seeks to create a Virtual Mouse programme that can be used in a variety of contexts and on a varietyof surfaces.

The following are the objectives of the project:

Design for mouse operation with the aid of a webcam. The Virtual Mouse technology works with the help of a webcam, which takes real-time photos and photographs. A webcam is required for the application to function.

The cursor is assigned to a certain screen position when the hand gesture/motion is converted into a mouse operation. The Virtual Mouse application is set up to identify the position of the mouse pointers by detectingthe position of the fingertips and knuckles on a defined hand color and texture.

Develop a multi user independent speech recognition system that captures voice in real-time with the help of a microphone and is able to retrieve folders, sub-folders, documents, copy, paste, left click, right click and double click by taking voice command and checking its validity.

Create a voice-activated mouse system that works in tandem with the gesture-activated system.

The most efficient and expressive way of human communication is through hand gestures and speech, which is universally accepted for communication. It is expressive enough for a dumb and deaf people to understand it. In this work, a real-world gesture system is proposed. Experimental setup of the system uses fixed position cost- effective web cam for high definition recording feature mounted on the top of the monitor of a computer or a fixed laptop camera. In addition to this, it uses a microphone to capture sound which is later processed to perform various mouse functions. Recognition and the interpretation of sign language or speech is one of the major issues for the communication with dumb and deaf people.

Python computer programming language has been used in the given project for the code, whereas OpenCV is used for computer vision to capture gestures. For hand tracking, the model in the proposed Virtual mouse system uses the MediaPipe package. The Python package Speech Recognition is used for voice instructions.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 14, May 2023

## II. RELATED WORK

In 1990, an early hardware-based system was introduced by Quam; in this system, it was mandatory for the user to wear a Data Glove.

Zheng you et al. (2001), proposed an interface system named Visual Panel that utilize arbitrary quadrangle-shaped planar object as a panel to allow the user to use any tip-pointer tools.

Kamran Niyazi et al. (2012), proposed color tracking mouse stimulation. . The said system tracks colour tapes on the user fingers by utilizing the computer vision technology.

Kazim Sekeroglu (2010), the system requires three fingers with three color pointers to simulate the click events.

Chu-Feng Lien (2015), proposed method that requires only finger-tips to control the mouse cursor and click events. The proposed system doesn't requires hand-gestures nor colour tracking in order to interact with the system, instead it utilize a feature name Motion History Images(MHI).

S. Shriram(2021);the model makes use of the MediaPipe package for the tracking of the hands and for tracking of the tip of the hands.

## III. LITERATURE SURVEY

**What is Human-Computer Interaction (HCI)?**

The study of how humans (users) interact with computers is known as human- computer interaction (HCI). It is a multidisciplinary field that deals with the design of computer technology. HCI began with computers and has now grown to embrace practically all aspects of information technology design.

**The Meteoric Rise of HCI**

When personal computers first became popular in the 1980s, HCI emerged at the same time as machines like the IBM PC 5150, Commodore 64, and Apple Macintosh began to be utilized in homes and offices. For the first time, sophisticated electronic systems such as games units, word processors and accounting aids were available to general consumers for use. As a result, as computers grew in size to the point where they were room-sized, expensive tools created exclusively for professionals in specialized situations, the necessity to research human-computer interaction that was also efficient and simple for less experienced users grew in importance. Design, computer science, psychology, cognitive science, and human- factors engineering are just a few of the fields that have been incorporated into HCI.

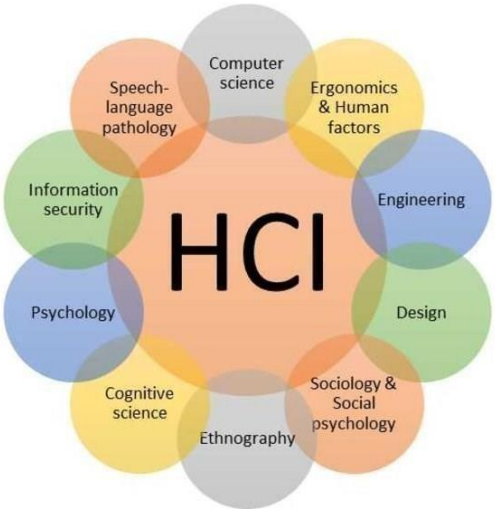



Fig.1

Throughout the research on human-computer interaction, several variants of these algorithms have been developed in various fields including Engineering, Design, social psychology, computer science , cognitive science , information security , sociology , and speech-language pathology are some of the fields covered.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 14, May 2023

The current research aims to create algorithms that lessen human reliance on hardware and strive for a more natural method of interacting with computers through hand gestures and speech.

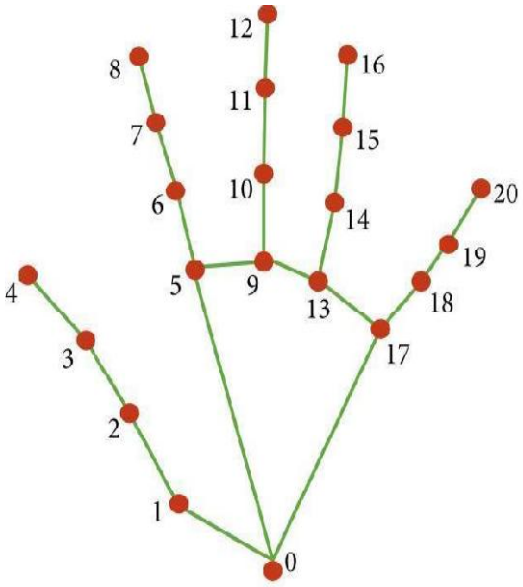
The OpenCV library is utilized for computer vision, while the MediaPipe framework is used to recognize and monitor hand motions. The system also employs machine learning techniques to track and recognize hand gestures and tips.

## IV. MEDIAPIPE

MediaPipe is a Google open-source framework that is used to apply in a machine learning pipeline. Because the MediaPipe framework is based on time series data, it issuitable for cross- platform development. The MediaPipe is a multimodal architecture that can be used with a variety of audio and video formats. The MediaPipe framework is used by developers to create and analyze systems using graphs, as well as to create systems for application development. The steps in a MediaPipe-enabled environment are carried out in the pipeline setup. The pipeline is scalable and runs on a range of platforms, including PCs, laptops, and mobile devices.

Performance evaluation, a framework for accessing sensor data, and a reusable collection of components known as calculators are the three primary components of the MediaPipe system.

In order to recognize and detect a hand or palm in real time, a single-shot detector model is used. The single-shot detector is used by MediaPipe. It is first trained for a palm detection model of hands in the hand detection module since palms of hands are easier to train and map. Furthermore, the non-maximum suppression is far more effective on small objects like hands and fists. The location of 21 joint or knuckle co- ordinates in the hand region makes up a model of hand map or landmark.




0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER_PIP
11. MIDDLE_FING
12. MIDDLE_FING
13. RING_FINGER_
14. RING_FINGER_
15. RING_FINGER_
16. RING_FINGER_
17. PINKY_MCP
18. PINKY_PIP
19. PINKY_DIP
20. PINKY_TIP

Fig.2

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 14, May 2023

## V. OPENCV

OpenCV is a real-time computer vision library that focuses on computer vision. Intel was the first to develop it. Under the open-source BSD license, the library is cross-platform and free to use. OpenCV is written in C++ and has a C++-based user interface. Python, Java, and MATLAB/OCTAVE all have bindings.

Open-source library of computer vision, image analysis, and machine learning. To do this, it has an infinity of algorithms that allow, just by writing a few lines of code, identifying faces, recognizing objects, classifying them, detecting hand movements.

## VI. PAST RESEARCHES

We have come a long way in the field of human computer interaction. Gesture based mouse control was carried out by wearing gloves initially. Later, colour

tips were also used for gesture recognition. Although such systems were not very accurate. The recognition accuracy is less due to use of gloves. Some users may not feel comfortable wearing gloves, and in some situations, recognition is not as accurate as it may be due to colour tip detection failure. Computer-based gesture detection systems have recently received some attention.

In 1990, Quam introduced a hardware-based approach that required the user to wear a DataGlove. Despite the fact that Quam's proposed method generates more precise results, certain of the gesture controls are difficult to execute with the system.

Zhengyou et al. proposed the Visual Panel interface system (2001). A quadrangle- shaped plane is used in this system, allowing the user to perform mouse operations with any tip-pointed interface instrument. Though the system can be operated contact free yet it does not solve the problem of surface area requirement and material handling.

Color tracking mouse stimulation was proposed by Kamran Niyazi et al. (2012). Using computer vision technology, the system tracks two colour tapes on the user's fingertips. One of the tapes will be used to control the cursor's movement, while the other will act as a trigger for the mouse's click events. Despite the fact that the proposed system handled the bulk of the issues, it only has a limited range of capabilities, as it can only perform fundamental actions such as cursor movements, left/right clicks, and double clicks.

To replicate click events, the system requires three fingers with three colour pointers, according to Kazim Sekeroglu (2010). The suggested system can detect pointers using colour information, track their motion, change the cursor according to the position of the pointer, and simulate single and double left or right mouse click events.

Chu-Feng Lien (2015) proposed a way for controlling the mouse cursor and click events using only one's fingertip. To interact with the system, the suggested system does not require hand motions or colour tracking; instead, it uses a feature called Motion History Images (MHI). Because the frame rates can't keep up with quickly moving objects, the proposed system can't detect them. Furthermore, because mouse click events occur when the finger is held in particular positions, this may cause the user to move their fingers constantly to avoid false alarms, which can be inconvenient.

S. Shriram(2021); the model employs the MediaPipe package for tracking the hands and the tips of the hands, as well as the Pynput, Autopy, and PyAutoGUI packages for moving around the computer's window screen and performing actions like left click, right click, and scrolling.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 14, May 2023

## VII. METHODOLOGY

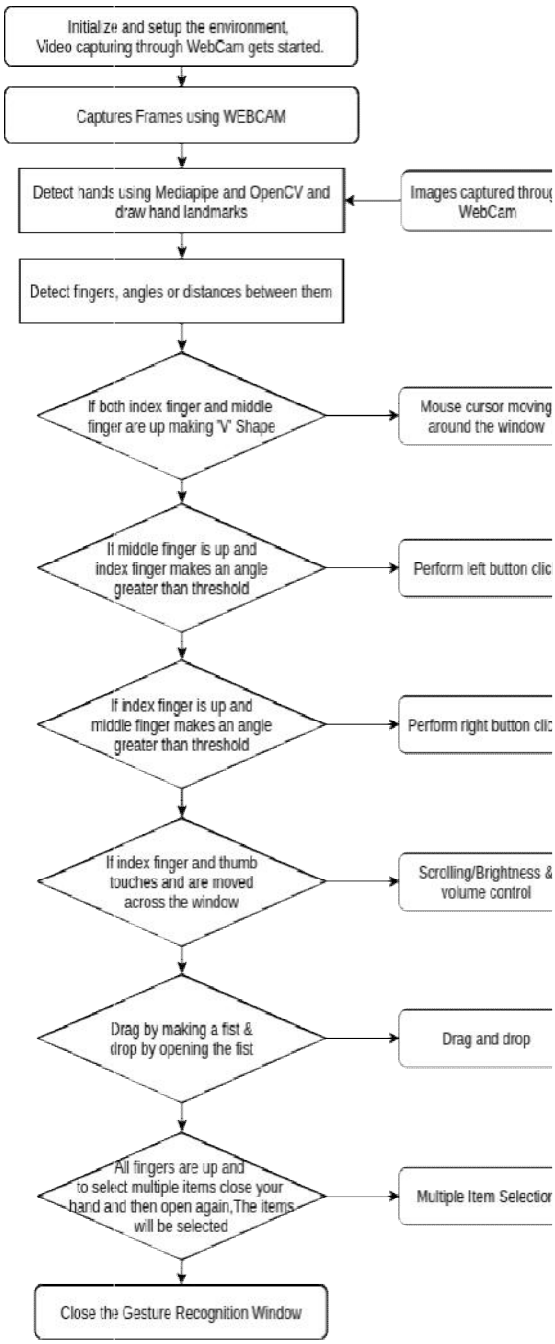



Fig 3

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 14, May 2023

**Voice Control**

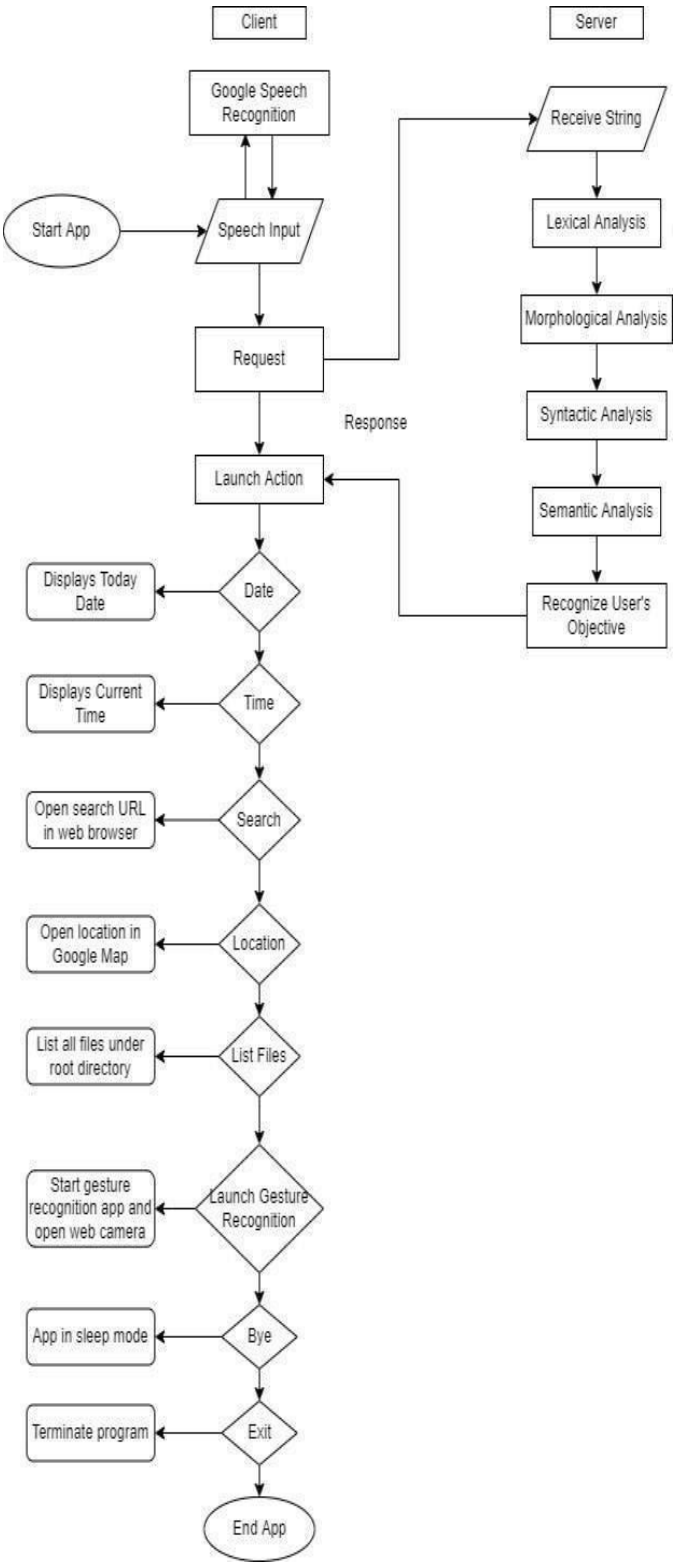



Fig.4

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 14, May 2023

## VIII. APPLICATIONS

Many applications benefit from the virtual mouse mechanism. It can be utilised to save space by eliminating the need for a physical mouse, as well as in instances where the physical mouse is not available. This technology reduces the need for a hardware device (mouse) and enhances human-computer interaction.

**Major applications:**

- Can alleviate physical stress on the body, which causes back discomfort, poorvision, and poor posture, among other things.
- Because it is not safe to use equipment bytouching them during the COVID-19 outbreak because contacting the gadgets could result in the virus spreading, the proposed virtual mouse can be utilised to manage the computerwithout using the physical mouse.
- Without the need of gadgets, this system can control automation systems androbots.
- Hand movements can be used to draw 2D images on the virtual system.
- Without the usage of a wireless or cable mouse, a virtual mouse can beutilised to play augmented reality and virtual reality games.
- This virtual mouse can be used by those who have difficulty with their handsto handle computer mouse functionalities.
- Using a combination of gesture and voice control, you can perform thefunctions of a traditional mouse quicklyand efficiently**.**

## IX. FUTURE SCOPE

Virtual Mouse will be introduced soon to replace the conventional computer mouse, making it easier for users to connect with and administer their computers. In order to correctly track the user's gesture, the software must be fast enough to capture and process every image and speech command.

Other features and improvements could be added to make the application moreuser- friendly, accurate, and adaptable in different contexts. The following are the enhancements and functionalities that are required:

**Smart Recognition Algorithm**

Using the palm and numerous fingers, additional functions such as enlarging and reducing the window, and so on, can be implemented.

**Better Performance**

The response time is largely influenced by the machine's hardware, which includes the processor's processing speed, the amount of RAM available, and the webcam's characteristics. As a result, when the software is performed on a respectable machine with a webcam that operates well in various lighting conditions and a better quality microphone that can detect voice instructions correctly and rapidly, the programme may perform better.

## X. ML PIPELINE (MEDIAPIPE) FOR HAND TRACKING AND GESTURE RECOGNITION

Mediapipe is a Machine Learning system built on the collaboration of pipeline models.

**What is ML Pipeline?**

A pipeline joins several stages together so that the output of one is used as the input for the next.

Pipeline makes it simple to train and test using the same preprocessing.

The hand tracking method makes use of a machine learning pipeline that consists of two models that work together:

- A palm detector that uses an aligned hand bounding box to locate palms on a whole input image.
- A hand landmark model that uses the palm detector's clipped hand bounding box to produce high- fidelity results. landmarks in 2.5D

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 14, May 2023

The following is a summary of the pipeline:

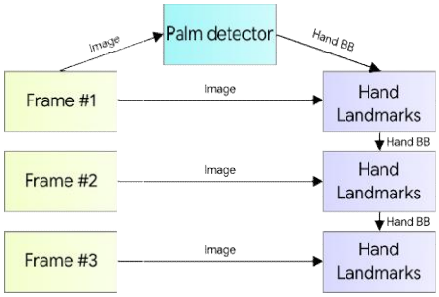



Fig.5

### 10.2. Palm Detector Model

Hand detection is a tedious process as it requires identifying hands of various sizes, shapes, with deformities, etc. It is more complex than face detection as the contrast in features is far less than that in face.

We use palm detection model first as detecting palm or a fist is much easier than detecting a full hand with articulated fingers. Also palms are smaller therefore non suppression algorithm works better for it.

## XI. CONCLUSION

The basic goal of the virtual mouse system is to control the mouse cursor and complete activities without needing a physical mouse by using hand gestures and voice commands. This proposed system is created by using a webcam (or any built-in camera) that recognizes hand gestures and hand tip movement and processes these frames to perform the relevant mouse actions using the notion of speech recognition to quickly follow voice commands and perform mouse activities.

The model upon rigorous testing has come out to be highly accurate and sophisticatedshowing enormous improvements with respect to prior existing models. Since the proposed model has been tested for high sophistication, the virtual mouse can be used for real-time applications. Because the proposed mouse system may be operated digitally utilizing hand gestures and voice commands rather than the traditional physical mouse, it will be of more value in combating the propagation of viruses like COVID-19 in the current context.

It functions as a useful user interface and contains all mouse features. Research on advanced mathematical materials for image processing and investigating different hardware solutions has made possible more accurate hand detections. Not only this project shows the different gesture operations and voice commands that can be done by the users but it can also demonstrate the potential in simplifying user interactions with personal computers and hardware systems. Yet a major extension to this work could be to be able to work at a more complex background and compatible with different light conditions

## REFERENCES

[1]. Himanshu Bansal, Rijwan Khan, "A review paper on human computer interaction" International Journals of advanced research in Computer Science and Software Engineering, Volume 8, Issue 4, April 2018

[2]. N. Subhash Chandra, T. Venu, P. Srikanth, "A Real-Time Static & Dynamic Hand Gesture Recognition System" International Journal of Engineering Inventions Volume 4, Issue 12,August 2015

[3]. S. Shiriam, B. Nagaraj, J. Jaya, "Deep learning based real time AI Virtual Mouse system using computer vision to avoid COVID-19 spread", Hindawi Journal of Healthcare Engineering, October 2021.

[4]. Hritik Joshi, Nitin Waybhase, Ratnesh Litoria, "Towards controlling mouse through hand gestures: A novel and efficient approach", Medi- caps University, May 2022

[5]. Mohhamad Rafi, Khan Sohail, Shaikh Huda, "Control mouse and computer system using voice commands", International Journal of Research in Engineering and Technology", Volume 5, Issue 3, March 2016

[6]. Li, M., Cheng, S., Li, H., & Wang, Y. (2018). A Voice-Controlled Human- Computer Interface Based on Deep Learning. IEEE Access, 6, 49978- 49987.

[7]. Pham, Q., Tran, D., & Tran, M. (2019). Vision- Based Gesture Recognition for Human-Computer Interaction: A Review. IEEE Access, 7, 13133- 13152.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 14, May 2023

[8]. Shrestha, P., Sallam, K., & Buratti, C. (2018). Hand Gesture Recognition Techniques: A Comprehensive Review. Journal of Ambient Intelligence and Humanized Computing, 9(4), 981-1016.
[9]. Shah, S., Jain, R., & Bharti, S. (2019). Hand Gesture Recognition using Deep Learning: A Review. Proceedings of the International Conference on Intelligent Sustainable Systems, 1- 5.
[10]. Oladeinde, O. A., Adetiba, E., & Sanusi, M. A. (2018). Design and Development of a Voice- Controlled Computer System Using Arduino. Journal of Information Technology and Software Engineering, 8(1), 1-7.
[11]. Zhang, Y., Yang, J., Yang, J., & Li, B. (2018). Gesture Control System Based on Accelerometer and Angular Velocity Sensor. Proceedings of the 2018 International Conference on Advanced Mechatronic Systems, 208-211.
[12]. Wang, T., Sun, Y., & Zhang, J. (2019). A Robust Vision-Based Gesture Recognition System for Human-Computer Interaction. Journal of Visual Communication and Image Representation, 59, 194-203.
[13]. Omondi, F. R., Kinyanjui, C. K., & Cheruiyot, A. K. (2020). Voice Control System Using Google Assistant and Arduino. International Journal of Engineering and Advanced Technology, 9(2), 183-188.
[14]. Tian, Y., Shi, Z., Li, H., & Ji, X. (2019). A Vision-Based Hand Gesture Recognition System for Human-Computer Interaction. Sensors, 19(4), 844.
[15]. Ahmed, M., Ali, A., Saeed, M., & Aslam, M. S. (2020). Virtual Mouse for Hand Gesture Recognition Using Image Processing. Proceedings of the 2020 7th International Conference on Signal Processing and Integrated Networks (SPIN), 824- 829.
[16]. Sarode, N. V., & Kharde, V. S. (2020). Voice and Gesture Controlled Robot Using Raspberry Pi. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 6(4), 2002-2006.
[17]. Fan, Y., Li, Y., & Qiao, H. (2020). Vision-Based Gesture Recognition for Human- Computer Interaction. International Journal of Advanced Computer Science and Applications, 11(11), 385-389.
[18]. Shaikh, F. M., Shaikh, F. M., & Kondekar, P. N. (2020). Gesture and Voice Controlled Smart Wheelchair for Disabled People. International Journal of Advanced Research in Computer Science, 11(6), 224-228.