Design and Implementation of an AI Virtual Mouse Using Hand Gesture Recognition

Dr. M Sharada Varalakshmi[1], A. Virinchi Sai[2], Chintagari Archana[3], Aliya Fatima[4]

[1] Professor, Methodist College of Engineering and Technology/Computer Science and Engineering, Hyderabad, India
[2,3] Artificial Intelligence and Data Science, Methodist College of Engineering and Technology, Hyderabad, India

Abstract: There have been many developments in the way human beings interact with computers, known as the field of Human-Computer Interaction (HCI) the most crucial being the Computer Mouse, aka that little plastic oval-shaped device that fits so nearly in our hands. As of now, the mouse still relies on other devices, it uses a battery for powering itself and to connect to the pc it uses a dongle. LEDs, switches, Touch screens, LCSs, Computer Mouse, etc., are commonly employed Human Machine Interface devices. Another way to interact with the devices is via hand gestures and this paper aims to overcome the limitation of the mouse to produce a recognition-based cursor control system using hand gestures recognition. The hand gestures are captured via the built-in cameras of the laptop/desktop or a webcam. The system or interface allows the user to control the cursor and perform right-click,left-click, scrolling functions, and resizing windows without the usage of a physical mouse, that way computers can be controlled virtually and can add to the notion of not relying on numerous devices such as mice, keyboards, joysticks, and so on to control laptop's functionalities.

Key words: Computer Vision, Human Computer Interaction (HCI), Image Processing, Hand Gesture Recognition, Human Machine Interaction (HMI).

1 Introduction

As technology is advancing at a rapid pace, devices that we utilize in our regular day-to-day life are becoming close-packed in the form of Bluetooth or other Wireless technologies. This paper takes a different approach to the general usage of a mouse the one that doesn't require the need for a mouse. This can be implemented by an Virtual Software that makes use of the concept called hand gesture recognition and detects hand tips for performing mouse functionalities. Hand detection technology is not new and has been used in the industry for a long time, such as in the fields of automation, IT Hubs, Banking Sectors, Medical Sciences, and so on. The primary motive for proposing a Virtual Mouse is to use the web cameras or the built-in cameras to interact with the computer to execute cursor functions, such as scrolling, etc.

The concept of using hardware or software technology to communicate with the machine is regarded as Human Machine Interaction or HMI. Over the years, several special input and output devices have been designed and developed with the sole purpose to ease the communication between humans and computers.

And with every little development in any device, is an attempt at making computers more intelligent and equipping the users with the ability to perform complex tasks on a computer. And this has been achieved by far, because of the continuous efforts made by computer scientists to successfully design Human-Computer Interfaces that met all the requirements for a general day-to-day action performed on a computer.

Vision and Sound are the main ways via which humans communicate, therefore having a Human-Machine Interface that uses vision and audio can be more intuitive and of greater use. Also having a visual system is helpful, as users can perform actions from a distance as well.

1.1 Problem Description and Overview

To design an alternate system for cursor control besides a traditional mouse. This system will be using the hand gesture recognition technology to perform operations of a mouse. This proposed AI virtual mouse will help overcome certain problems that comes with using a traditional mouse, such as when there's not enough space to operate via the mouse, or if a person with disabilities is not able to control it properly. This can also be helpful during some pandemic or when Covid-19 struck, using a virtual mouse will reduce the spread of diseases that occurs via touching different day to day devices. Thus, this approach of a virtual mouse will solve these problems, as simple hand gesture recognition can be implemented via a webcam.

1.2 Objective

The foremost motive of developing a virtual mouse here is to provide an alternative to the traditional cursor control via a mouse. This cursor control actions of the mouse can be implemented with the help of hand gesture and hand tips recognition.

2 Related Work

Many attempts have been made for gesture recognition for Human-Computer Interaction, and the two possible ways to achieve this are hardware-based and via computer vision.

[1] One of the earliest proposals was made by D.L. Quam in his paper, "Gesture Recognition with DataGlove", where he used a DataGlove, which is an electronically-instrumented glove that helps in the identification of hand and finger positions. He investigated 22 gestures in total. Even though it showed high accuracy for controlling, it is not feasible as wearing a bulky electronic DataGlove is difficult. Also, it requires some particular set of sensors to accurately recognize a gesture from the group of gestures. [2] In 2010, "A real-time hand gesture recognition system using motion history image" was proposed by Chen-Chiung, wherein they used an adaptive skin color model to detect skin color regions like hands. It defined six different hand gestures i.e., up, down, left, right, fist hand & waving. It showed an accuracy of 94.1% and a processing time of 3.81. Although it is feasible it is more suitable to control most home appliances. It doesn't have the flexibility to go beyond those 6 gestures, which makes it difficult. [3] In 2011, the paper "A Human-Machine Interaction Technique: Hand Gesture Recognition Based on Hidden Markov Models with Trajectory of Hand Motion", used visual hand gesture recognition for cursor control. It uses a hidden Markov model classifier to effectively separate face and hand regions. In this system, gestures are of single or both hands. The authors had defined 8 gestures for HMI, the efficiency was very satisfactory but it works only on high-configuration systems. [4] In 2013, the paper "Real-Time Static and Dynamic Hand Gesture Recognition" suggested a real-time, static, and dynamic hand gesture recognition system that allows users to interact with systems intuitively and naturally, it included the process of image acquisition, hand segmentation, hand feature extraction, gesture recognition, etc. But this algorithm was unable to carry out the detection in complex or dark backgrounds. [6] In the paper "Cursor Control System Using Hand Gesture Recognition", they developed a machine-user interface to achieve hand gesture recognition by implementing simple computer vision and multimedia

techniques. [7] Abhik Banerjee, and Abhirup Ghosh in their 2014 paper, “Mouse Control using a Web Camera based on Color Detection’, proposed a different approach to HCI wherein hand gestures were acquired via the webcam based on color detection algorithms. According to their approach first, the image has to be converted into a greyscale, different colors are detected and extracted from that gray image, then converted to a binary image, calculating the centroid, simulating left & right click by using different color pointers. The drawback of this system is that due to the presence of other colored objects in the background, color detection algorithms might not work fairly well. This requires the background to be light and have no other bright-colored objects. This system will also run slower because of high computational requirements. [8] Yimin Zhou in his paper “A novel finger and hand pose estimation technique for real-time hand gesture recognition” used a high-level hand feature extraction method, he modeled fingers as cylindrical objects to distinguish fingers through hand edges. They developed a system that can not only extract extensional fingers but also fluxional fingers with high accuracy. This was a clever approach as hand rotation and finger angle variation did not affect the system’s performance. [9] Paul Robertson in his paper ‘Virtual Mouse Vision Based Interface’ developed a mouse interface which uses optical flow in a region and Kalman Filter for smooth movement of the kiosk mouse pointer. The system was trained on thousands of images of hand signs collected from their lab, which led to the higher accuracy of the visual recognition system. [10] Continuously Adaptive Mean Shift algorithm or Camshift algorithm was used to develop virtual mouse by R. O. Dogan, the study comprised of three essential steps: hand gesture tracking, features of hand region extraction and classification of the features.

3 Algorithms used for hand tips detection

For recognizing the hand gestures, hand tips and movement of hand, the framework called MediaPipe is used, and for machine vision OpenCV library is used.

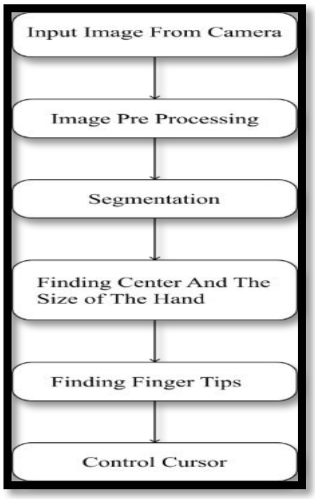

Fig. 1. Flow Chart of the proposed system

MediaPipe: It is an open-source cross-platform framework developed by Google that allows developers to build and deploy scalable machine learning pipelines and computer vision projects for various types of applications. It provides a wide range of pre-built, customizable components, including algorithms for face detection, hand tracking, pose estimation, and object detection, among others.

MediaPipe is designed to be a flexible and modular framework, allowing developers to customize and combine the different components to create a pipeline that suits their needs. It supports multiple programming languages, including, C++, Java, Python and runs on a variety of platforms, including desktops, mobile devices, and embedded systems.

One of the key features of MediaPipe is its ability to perform real-time, high-quality inference on live video streams, making it suitable for applications that require low-latency and high accuracy, such as augmented reality, virtual try-on, and gesture recognition. The framework also includes a range of tools for data preprocessing, model training, and performance analysis, which makes it easier for developers to build and deploy their computer vision and machine learning applications.

MediaPipe has been used in various research and industry applications, including Google's Live Caption, an accessibility feature that automatically captions live audio on mobile devices, and the AI-powered dance app, Tangiplay, which uses hand tracking and pose estimation to teach kids how to dance.

Overall, MediaPipe is a powerful and versatile framework that simplifies the development of computer vision and machine learning applications, making it an attractive option for developers and researchers alike.

Open CV: Open CV (Open-Source Computer Vision Library) is an open-source computer vision and machine learning library that allows developers to build applications that makes use image processing, object detection, hand gesture detection and recognition, and much more. It was first released in 2000 and has been continuously updated and improved since then. It provides an extensive set of functions and algorithms for processing and analyzing images and videos, including edge detection, feature detection, image segmentation, and machine learning-based object detection.

Open CV has been used in various research and industry applications, including face recognition, gesture recognition, and augmented reality. It is also used in many popular open-source projects, such as TensorFlow and PyTorch, for pre-processing and post-processing of data.

Overall, OpenCV is a powerful and flexible tool for developers and researchers working in the field of machine learning, computer vision and other trending technologies. It continues to evolve and improve, and its wide range of functionalities and platforms make it a valuable asset for many different applications.

AutoPy: AutoPy is a free, open-source cross-platform desktop automation tool that enables developers to automate various tasks on their computer. It provides a simple, easy-to-use API that allows developers to simulate keyboard and mouse input, capture and manipulate screen images, and perform window management functions.

This allows developers to automate tasks that require user interaction, such as clicking buttons, filling out web forms, and navigating through menus. AutoPy also provides functions for capturing and manipulating screen images, which can be useful for tasks such as image recognition, hand gesture recognition and OCR (Optical Character Recognition).

Another useful feature of AutoPy is its window management capabilities. Developers can use AutoPy to resize and move windows, hide or show windows, and perform other window-related tasks. This can be useful for automating tasks that involve multiple windows or applications.

4 Methodology

To implement the AI Virtual-Mouse we used the following procedure:

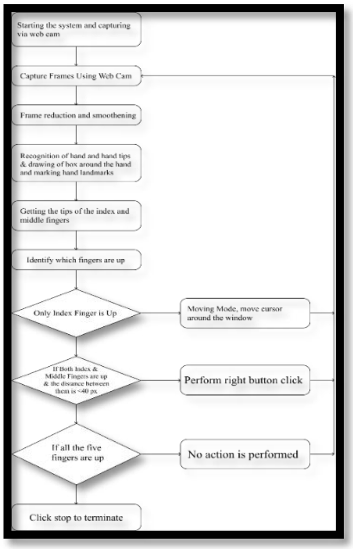
Starting the system and capturing via web cam
Capture Frames Using Web Cam
Frame reduction and smoothening
Recognition of hand and hand tips & drawing of box around the hand and marking hand landmarks
Getting the tips of the index and middle fingers
Identify which fingers are up
Only Index Finger is Up
Moving Mode, move cursor around the window
If Both Index & Middle Fingers are up & the distance between them is <40 px
Perform right button click
If all the five fingers are up
No action is performed
Click stop to terminate

Fig. 2. Process flow of the AI-Virtual Mouse

i) Initializing the system:

We used Visual Studio Code for creating this system, we initialized the working space by installing all the necessary libraries and frameworks (AutoPy, OpenCV, MediaPipe, &Numpy). We imported these libraries into our code and started writing the algorithms for the same.

ii) Video Capturing and processing

This system is uses frames to accurately capture the hands, using OpenCV we can capture the video, the web camera captures and passes to the hand recognition module. Each frame is then smoothened to accurately distinguish hand tips from other objects.

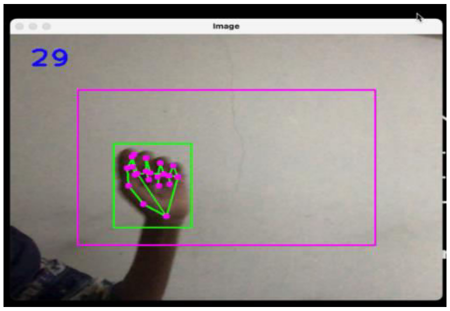

Image 1. Detection of hands and hand tips from the web cam

iii) Transforming to coordinates

The hand tips detected will be converted into coordinates to control full screen using computer window via different gestures. As shown in Image 1, a rectangular box is drawn when the hand is detected, it's drawn across the window through which we can move the cursor over the whole region.

iv) Identifying which fingers are up

During this step, we detect which fingers are up and depending on that finger and its coordinates we implement the respective cursor functionality.

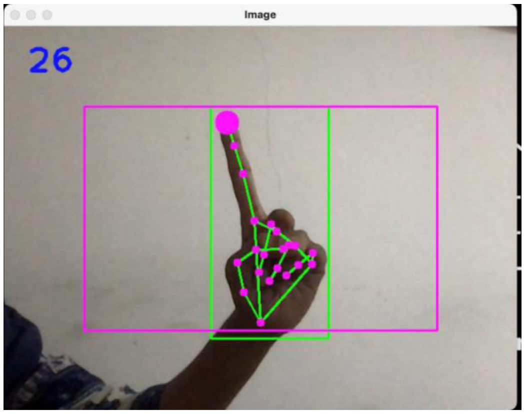

Image 2. Determining which fingers are up

a) Moving Mode

When the finger tips are detected and when the finger with the tip Id =1or only the index finger is up then the mouse is in moving mode, i.e., it navigates through the window. This is achieved via AutoPy.

b) Clicking Mode

When the finger tips with id =1 & id = 2 are up, i.e., when index finger and middle finger are up, the and the distance is below 40 px between them, then the mouse is in clicking mode, right clicking mode. As shown in Image 4 & 5.

c) No action is performed

No action is performed when all fingers are up, as shown in Image 3.

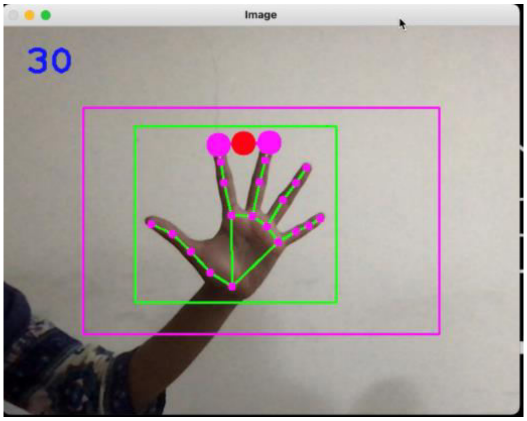

Image 3. All five fingers are up, no action being performed

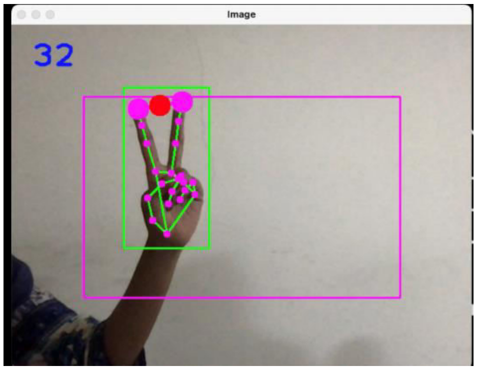

Image 4. Middle and index fingers are up, but distance is > 40px

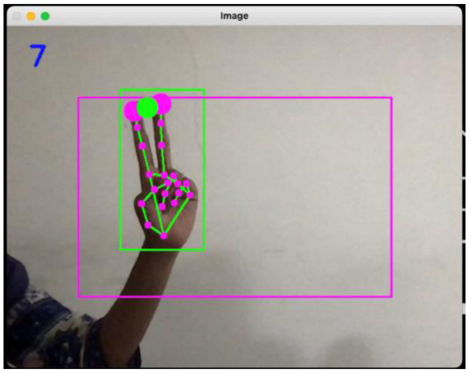

Image 5. Middle & Index fingers are up, & distance is < 40 px mouse is in clicking mode

5 Future scope

The AI-Virtual mouse system proved to be effective due to its high refresh rate and the ability to open and close folders and navigate through them using hand gestures. However, there is room for further improvement. One way to enhance the system would be to add more gestures, providing users with greater flexibility.

Additionally, we could include customization options that allow users to choose which fingers to use for clicking different applications and navigating the computer.

To achieve these improvements, we would need to enhance the hand gesture recognition algorithms significantly. This would require using high-quality web cameras or high-definition web cameras that are capable of capturing clear and detailed images. Furthermore, a large amount of data containing different skin colors and hand orientations would need to be fed into large machine learning models to train them effectively. This would increase the computational complexity required to process images and train the models.

By incorporating these changes, the limitations of the AI-Virtual mouse system could be significantly reduced. While the system was developed on a Macintosh OS computer, it may perform even better on other operating systems, such as Linux or Windows.

7 Applications

An AI Virtual Mouse system has a wide range of applications across various industries, a few of them are:

- Accessibility: An AI-Virtual Mouse system can be utilized to provide access to computers for physically handicapped people, allowing them to interact with their computers using hand gestures instead of a traditional mouse system.
- Education: This system can be used in educational settings to help students with special needs interact with computers and learn more effectively.
- Medical: AI Virtual mouse systems can be used in medical applications to control medical equipment using hand gestures, reducing the risk of contamination and infection.
- Industrial automation: AI Virtual mouse systems can be used to control machines and equipment using hand gestures, reducing the need for physical buttons and switches.
- Electricity: Having to control things with gestures reduces the risk of shocks and other accidents.
- Gaming: This system would provide a more immersive experience for gamers, as they could gain more control over the game characters and game menus.

If implemented rightfully and integrated with different customizations and by the usage of high-fidelity cameras an AI Virtual mouse system has the capacity to be used in the place of traditional mouse and can revolutionize the way we interact with computers and other devices, opening up new possibilities for a wide range of innovations and developments.

7 Conclusion

The main goal of this project is to use hand gesture technology to create a virtual mouse wherein we can operate the mouse using hand gestures.

In this paper, we tied to create an AI virtual mouse to operate the computer without the physical mouse. We can see the movement of the mouse on the screen via hand gestures. Additionally, the closing and opening of the files can be done through hand gestures only. To track hand gestures for this project, a camera will be needed. To function effectively, the user's hand must be in front of the camera. The camera will be utilized to track the hand's motion and translate it into on-screen cursor movement. visual computing For hand tracking, Python tools like Autopy, OpenCV, and MediaPipe will be used. The algorithms will first track the hand motions and then translate them into the appropriate screen activities.

References

1. D. L. Quam, G. W. Williams, P. C. Browne and J. R. Agnew.: An Experimental Determination of Human Hand Accuracy with a DataGlove. In Proceedings of the Human Factors Society 33nd Annual Meeting, pp. 91-95.(1989)
2. Chen-Chuiung, D. Liou, David Lee.: A real time hand gesture recognition system using motion history image. 2nd International Conference on Signal Processing Systems (2010)
3. Chang-Yi Kao and Chin-Shyurng Fahn.: A Human-Machine Interaction Technique: Hand Gesture Recognition Based on Hidden Markov Models with Trajectory of Hand Motion.In Procedia Engineering (2011)
4. P.S. Neethu.: Real Time Static & Dynamic Hand Gesture Recognition. In International Journal of Scientific & Engineering Research, March 2013
5. K. H. Shibly, S. Kumar Dey, M. A. Islam and S. Iftekhar Showrav.: Design and Development of Hand Gesture Based Virtual Mouse. 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh.(2019)
6. Ashwini M. Patil, Sneha U. Dudhane, Monika B. Gandhi, and Nilesh J. Uke.: Cursor Control System Using Hand Gesture Recognition. In International Journal of Advanced Research in Computer and Communication Engineering, Vol .2, Issue 5, May 2012
7. Abhik Banerjee, Abhirup Ghosh, Koustuvmoni Bharadwaj, & Hemanta Saika.: Mouse Control Using a Web Camera Based on Color Detection.In International Journal of Computer Trends and Technology (UJCTT) – Volume 9 number 1- Mar 2014
8. Yimin Zhou, Guolai Jiang, Yaorong Lin.: A novel finger and hand pose estimation technique for real time hand gesture recognition.
9. Paul Robertson, Robert Laddaga, and Max Van Kleek.: Virtual Mouse Vision Based Interface. In International Conference on Intelligent User Interfaces, 2004,
10. R. O. Dogan, H. Dogan and C. Kose.: Virtual mouse control with hand gesture information extraction and tracking. In 23rd Signal Processing and Communications Applications Conference (SIU), Malatya, Turkey, 2015, pp. 1893-1896.