# DESIGN AND IMPLEMENTATION OF AI VIRTUAL MOUSE AND ASSISTANT

Kapil Gupta
Computer Science & Engineering
(2021494203.kapil@ug.sharda.ac.in)
Sharda University,
Greater Noida (India)

Tushar Pandey
Computer Science & Engineering
(2021826063.tushar@ug.sharda.ac.in)
Sharda University,
Greater Noida (India)

***Abstract*— In this work, a virtual mouse system is provided that uses a webcam to for:**

**record user movements and a voice assistant to navigate system controls fast. The technology will allow the user to move the computer cursor with hand motions using Media Pipe. It will perform actions like left clicking and dragging using a variety of hand motions. You can also select a variety of options, and change the brightness and loudness. The system is constructed using advanced Python packages like Media Pipe, Pyautogui, OpenCV, etc. All activities are physically controlled by hand motion and voice assistance. The research uses cutting-edge machine learning and computer vision techniques, which operate well without the need for additional computing resources, to recognize hand movements and spoken instructions..**

## Technologies to be used:

HUMAN COMPUTER INTERACTION (HCI)

MOTION HISTORY IMAGES (MHI)

INTEGRATED DEVELOPMENT ENVIRONMENT (IDE)

Open CV (OPEN COMPUTER VISION)

NATURAL USER INTERFACE (NUI)

MEDIAPIPE

PautoGuI

## 1) INTRODUCTION

Gestures are used to communicate nonverbally and to conveying a certain message. This message can be sent through person's hand movements. When communicating with others, gestures can be used to communicate and convey a message individuals, from easy to highly difficult hand motions. For instance, we can use a number of easy gestures to motions that are expressed in sign languages that are integrated

with their syntax and diction, more often known as sign languages, to point to anything (an object or people), or use the sign language's syntax and lexicon. As a result, with the aid of computers, humans can interact with one another more efficiently by using hand gestures as a tool.

The movement of a visual object is one mouse function that has been replaced by hand movements. The task is designed to be inexpensive, and it makes use of inexpensive input devices like a webcam to record hand movements. To manipulate materials, preset command-based movements are simulated.

## 2) LITERATURE SURVEY

A. Recognition of Hand Gesture:

Gesture recognition is a hot topic in computer science and is involved in developing systems that translate human movements so that anyone can interact with a device without touching it. Gesture recognition is the process of recognizing, presenting, and converting gestures into precisely intended commands. The goal of hand gesture recognition is to identify certain unique hand gestures and process the gesture expressions on devices that use map as output.

From various sources, here are his three methods of recognizing hand gestures:

- Machine Learning Methods
- Algorithm Methods
- Rule-based Methods

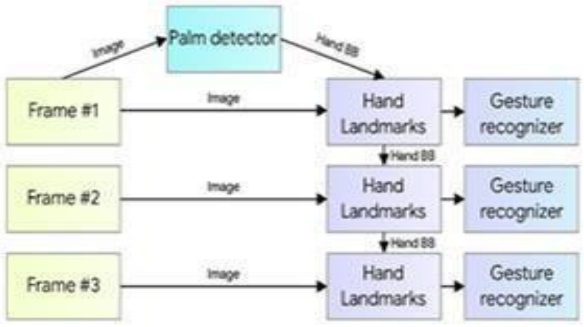


Fig1: Hand Perception Pipeline overview

A. Media Pipe Framework:

Media Pipe gives life into the products and services we use every day. Unlike other resource-intensive machine learning frameworks, Media Pipe uses very little. It is so small and powerful that it can run on embedded IOT Devices. It is Launched in 2019, Media Pipe opens up a whole new world of possibilities for researchers and Developers.

Hand Gesture recognition:

- Palm detector model
- Hand landmark model
- Gesture recognizer

**B. Voice Assistant**

Below figure forms the basis for any kind of voice assistant.

| VOICE TECHNOLOGY | BRAIN TECHNOLOGY |
|---|---|
| Voice Activation | Voice Biometrics |
| Automatic Speech Recognition (ASR) | Dialog Management |
| (Teach-To-Speech (TTS) | Natural Language Understanding (NLU) |
| | Named Entity Recognition NER) |

Fig2: Technologies for Voice Assistant

## 3) Technologies are Used:

1. HUMAN COMPUTER INTERACTION(HCI):

An interdisciplinary field of study called Human-Computer Interaction (HCI) is devoted to the design of computing technology, and more specifically to the interaction between humans (users) and computers. HCI initially focused primarily on computers, but has since expanded to include almost all aspects of IT design.

2. MOTION HISTORY IMAGES (MHI):

Images of motion history provide a static description of a temporal action. They describe the movement of an object or a subject over time, including the direction they are travelling in and the timing of their actions. We can produce helpful generalizations about the appearance of gestures by teaching a classifier with a database of actions. We can produce labels for new data based on an analysis of the new motion history image after we have our classifying network.

3. INTEGRATED DEVELOPMENT ENVIRONMENT (IDE):

An integrated development environment (IDE) is a piece of software that helps in the efficient development of software code for programmers. Combining functions like software editing, building, testing, and packaging in a user-friendly program, it improves developer productivity. Software developers use IDEs to make their jobs easier in the same way authors use text editors or accountants use spreadsheets.

4. OpenCV:

A computer vision and machine learning software library called OpenCV is available for free use. A standard infrastructure for computer vision applications was created with OpenCV in order to speed up the incorporation of artificial intelligence into products. OpenCV makes it simple for businesses to use and alter the code because it is a product with an Apache 2 license.

5. NATURAL USER INTERFACE (NUI):

A user interface that aims to feel as natural to the user as possible is known as a NUI. An NUI aims to eliminate the appearance of the interface by enabling seamless interaction between humans and machines. A touchscreen interface is a typical illustration of a natural user interface since it enables you to move and control objects by tapping and dragging your finger(s) on the screen. Your touch has an effect on the digital items on the screen in a manner similar to how physical objects would. Compared to using a keyboard and mouse to interact with the objects on the screen, employing a touchscreen interface's direct feedback feels more natural.

6. MEDIAPIPE:

MediaPipe offers cross-platform, customizable ML solutions for live and streaming media. End-to-End acceleration: Built-in fast ML inference and processing accelerated even on common hardware. Build once, deploy anywhere: Integrated solutions

work across Android, iOS, desktop/cloud, web, and IoT..

## 4) METHODOLOGY:

The proposed system can be activated initially using either a gesture control software or a voice assistant programme. The other programme can also be launched with either. Users' gestures are recorded using a webcam for the gesture control software, and each frame is run via MediaPipe's hand gesture recognition module (mp.solutions.hands) to identify landmarks. With the aid of some computation, a gesture is detected using these landmarks. A controller class then executes these commands using actions. This is repeated frequently. The voice assistant programme uses a microphone to record user input. Commands are recognised. Actions are carried out in accordance with the commands.

The project uses touch control to provide the following functions:

1) Move the cursor
2) Stop gesture
3) Left cursor
4) Double click
5) Scrolling
6) Drag and Drop
7) Multiple Item Selection
8) Volume Control

The project uses voice assistant to provide following functions:

a) Launch / Stop gesture recognition
b) Google Search
c) Find a location on Google maps
d) File navigation
e) Date & time
f) Copy Paste
g) Sleep/wake voice assistant
h) Exit

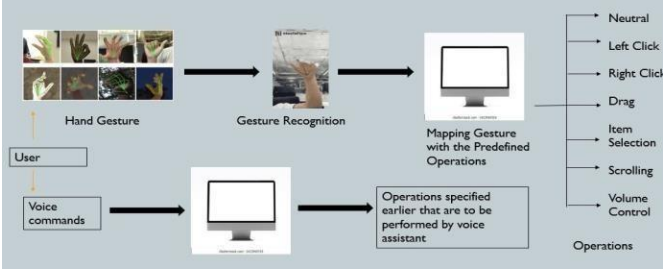


Fig3: Architecture Diagram of Proposed System

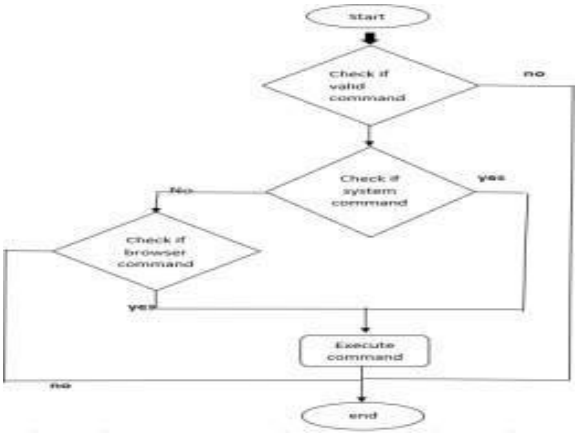


Fig4: Flowchart of Voice Assistant

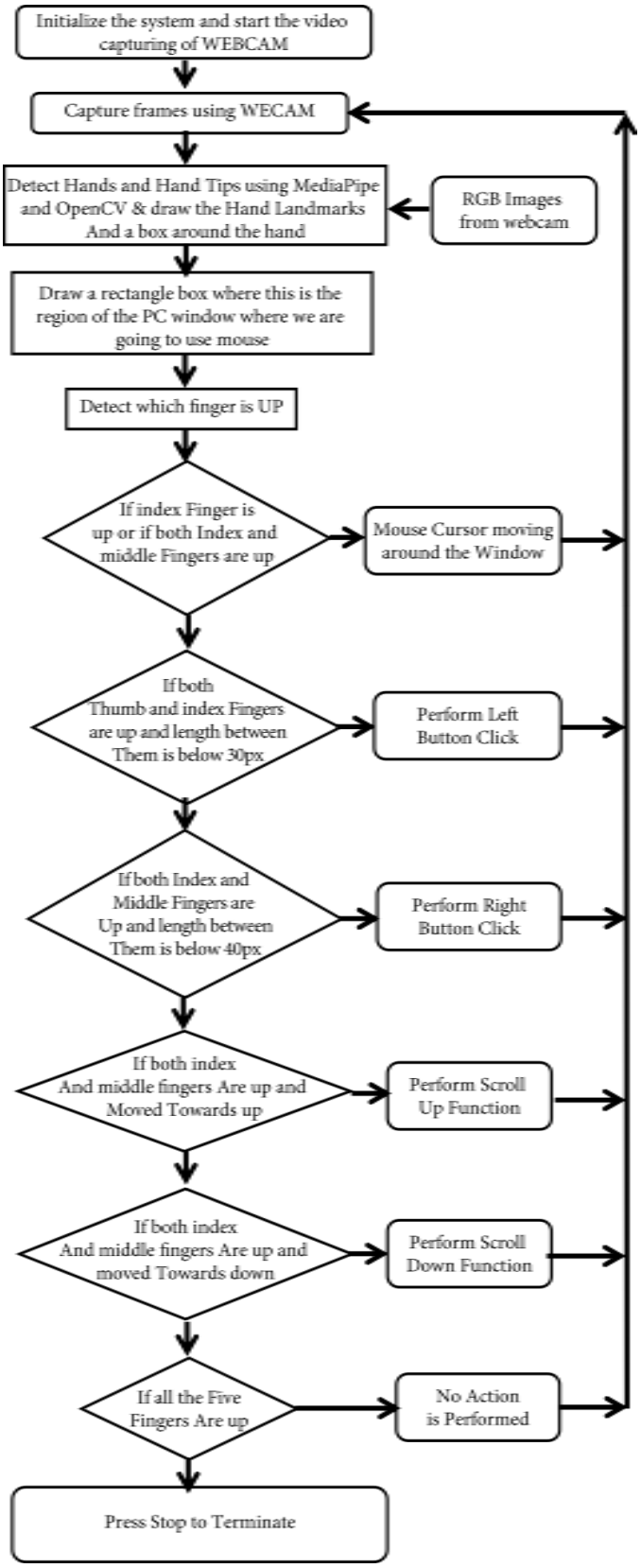


Fig5: Flowchart of virtual mouse

## 5) Experiment Result:


**Fig4 : Hand detected by Webcam**


**Fig5: Connected with mouse**

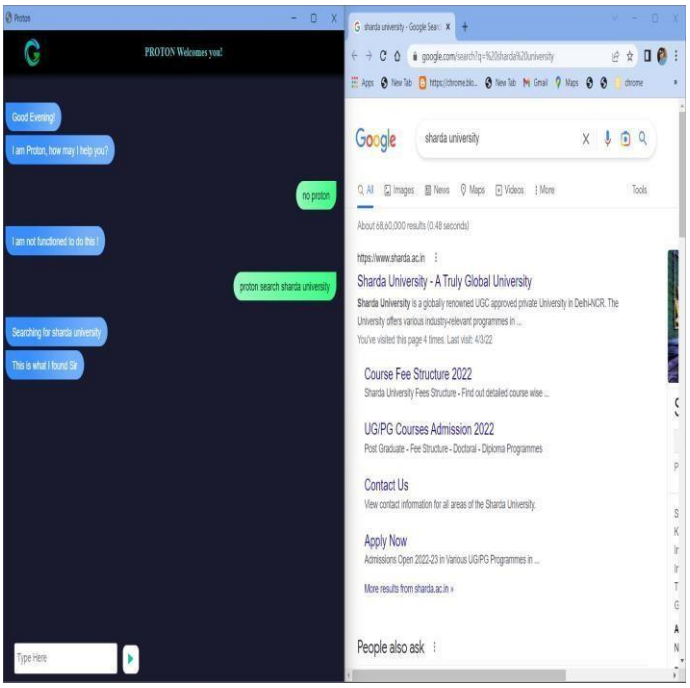


**Fig6: connected with Voice Assistant**

## 6) FUTURE SCOPE OF PROJECT:

Our objective is to provide more gestures so that users can eventually complete more tasks efficiently. This proposal proposes a system that uses only of the proper right hand when making gestures. As a result, future improvements to the current technique in use will allow us to use the of both hands for variety of gestures. The quick advancement of hand gesture recognition technologies has greatly enhanced numerous applications. It also enables online learning and distance training, can operate televisions, and helps doctors and surgeons examine each patient's images more vividly. One of the best applications is signing language interpretation for people who have hearing loss The project must achieve the objectives for which it was created and demonstrate effectiveness in execution.. And we've automated activities like locating places on Google Maps, navigating documents, starting and stopping gesture detection, doing Google searches, and dozing off and waking up the voice assistant through the voice assistant function. The end user will save time and effort thanks to this functionality, which will also make computers more accessible to people who are blind or disabled.

## Conclusion:

In this project, we are developing a method for using a live camera to control the mouse pointer. The system can perform all Mouse tasks and is based on computer vision techniques. However, because there are so many different human races, it is challenging to obtain solid results. Utilizing this technique would make presentations simpler and save workspace. Using the Palm and numerous fingers, it offers functionality like window closing, window enlargement, and more. This software is made to make physically challenged persons use desktops and laptops as intelligently as regular people do, taking into account the utilization of new cutting-edge technologies. This project was created using Python 3.8 (64-bit) and open source modules, making it ideal for improvements in the future.

## Acknowledgement:

We would like to express our sincere gratitude to our mentor Mr. Dharmraj Sharda university, Greater Noida for his guidance, encouragement and valuable suggestions.

## REFERENCE:


- https://www.topcoder.com/thrive/articles/what-is-the-opencv-library-and-why-do-we-need-to-know-about-it
- https://dx.doi.org/10.2991/aer.k.211106.017
- https://github.com/nicknochnack/MediaPipeHandPose/blob/main/Handpose%20Tutorial.ipynb
- https://www.youtube.com/watch?v=9iEPzbG-xLE
- https://learnopencv.com/introduction-to-mediapipe/

- https://ijirt.org/master/publishedpaper/IJIRT152099_PAPER.pdf
- https://ijirt.org/master/publishedpaper/IJIRT155856_PAPER.pdf