PROJECT BASED LEARNING (PBL-1) LAB (CSP254)

AI VIRTUAL MOUSE AND VOICE ASSISTANCE

B.TECH 2nd YEAR

SEMESTER: 3rd

SESSION: 2022-2023

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

SHARDA SCHOOL OF ENGINEERING & TECHNOLOGY

SHARDA UNIVERSITY, GREATER NOIDA

Submitted By:

Name: Kapil Gupta , Tushar Pandey

Roll No: 210101237,210101534
System id: 2021494203,2021826063
Section: J

Submitted To: Mr.Dharmraj

Table of Contents

Project Title: **VIRTUAL MOUSE AND VOICE ASSISTANCE**

Team / Group Formation:

S. No	Student Name	Roll Number	System ID	Role
1	Kapil Gupta	210101237	2021494203	Code, Research paper
2	Tushar Pandey	210101534	2021826063	Code, Report

Technologies to be used

HUMAN COMPUTER INTERACTION (HCI)

MOTION HISTORY IMAGES(MHI)

INTEGRATED DEVELOPMENT ENVIRONMENT(IDE)

OpenCV (OPEN COMPUTER VISION)

NATURAL USER INTERFACE (NUI)

MEDIAPIPE

PYAUTOGUI

Problem Statement:

To develop a problem solution we first need to understand the problem that the user is facing difficulty while using the program. The solution of the existing problem is been eliminated in this. Our flow of work is to process.

1. By using open resources
2. no sensors use
3. Proper resource utilization
4. Low cost

Literature Survey:

A. Recognition of Hand Gestures

Gesture recognition is a topic in computer science topic that deals with developing systems that translate human movements so that anyone can interact with a device without actually 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 cards as output.

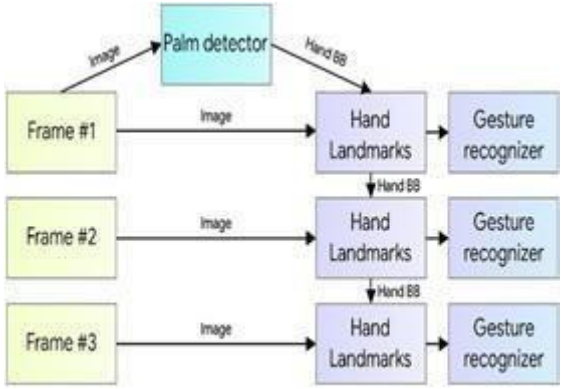


Fig. Hand perception pipeline overview

From a variety of sources, there are three ways to recognize hand gestures, as follows:

1. Machine Learning Methods
2. Algorithm Methods
3. Rule-based Methods

A. Media Pipe Framework

Media Pipe brings the products and services we use every day to life. Unlike other resource-intensive machine learning frameworks, We rarely use Media Pipe. It's 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 as follows:

1. Palm model detector
2. Hand model landmark
3. Gesture recognizer and detector

B. Voice Assistant(Proton)

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)

Fig. Technologies for Voice Assistant

Project Description

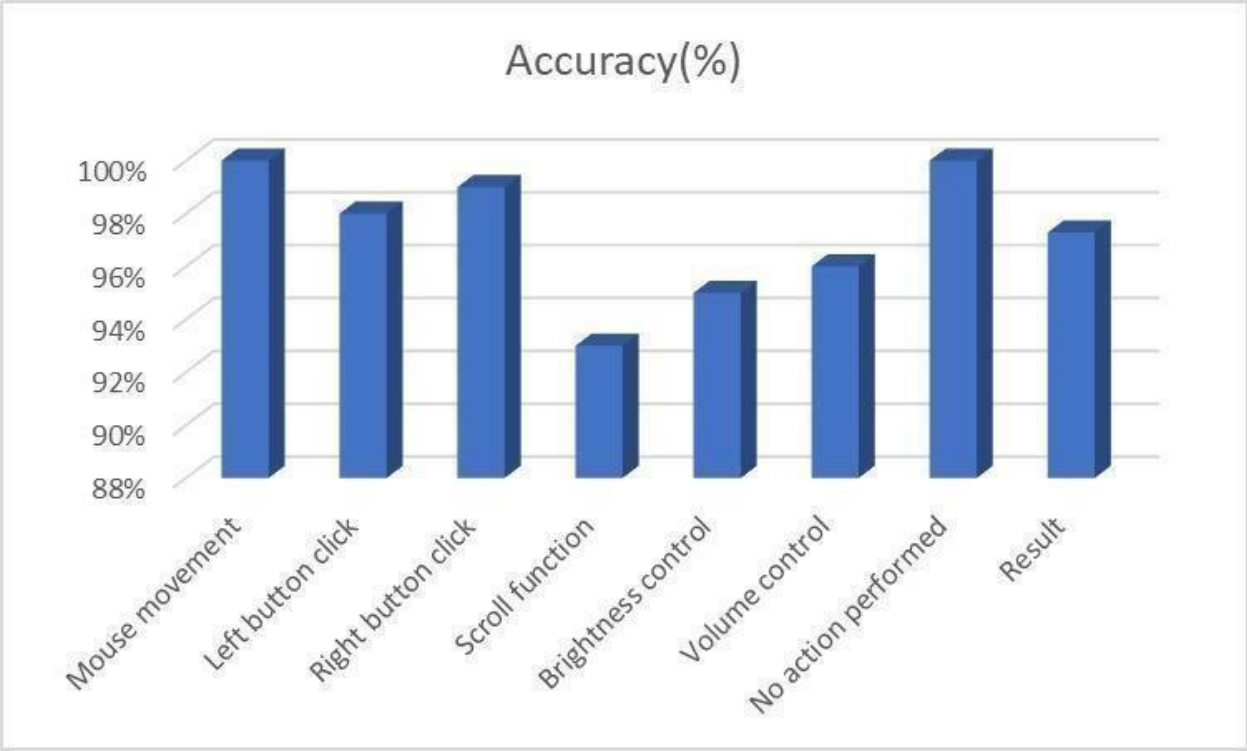
The proposed AI virtual mouse system, given the concept of using computer vision to facilitate human-computer interaction.

Inter-comparison of tests of AI virtual mouse systems is difficult due to the limited number of available datasets. Hand gestures and fingertip recognition were tested in different lighting conditions. It was also tested at different distances from the webcam to track hand gestures and fingertip recognition. Experimental tests were performed to summarize the results shown in Table.

The test was run 25 times by his 4 people and 600 gestures with manual labeling. The test was run in different lighting conditions and different distances from the screen, with each person testing her AI virtual mouse system ten times his. Test results are shown in the table under normal lighting conditions, 5 times in low light, 5 times at close distance from webcam, 5 times at far distance from webcam.

Mouse function performed	Successes	Failuree	Accuracy(%)
Mouse movement	100	0	100%
Left button click	98	2	98%
Right button click	99	1	99%
Scroll function	93	7	93%
Brightness control	95	5	95%
Volume control	96	4	96%
No action's perform	100	0	100%
Results	681	19	97.28%

From the table, we can see that the proposed AI virtual mouse system achieved an accuracy of about 97%. This 97% accuracy of the proposed AI virtual mouse system shows that the system worked well. As can be seen from the table, the "scrolling function" is the gesture that computers have the least understanding of, so it has the lowest accuracy. Scrolling functions are less precise because the gestures used to perform certain mouse functions are more difficult. Accuracy is also very good, high on all other gestures as well. Compared to previous virtual mouse approaches, our model performed very well with 97% accuracy. The accuracy chart is shown in the figure



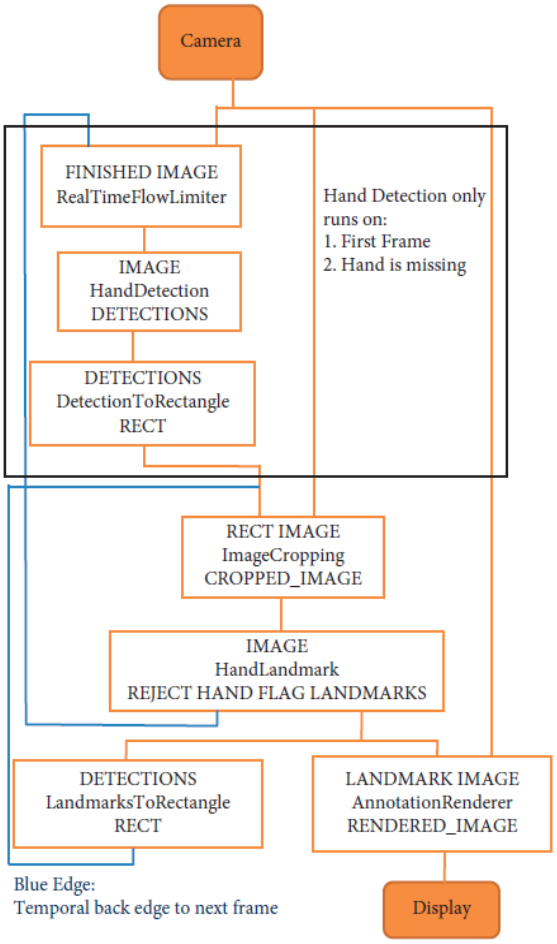
Project Modules: Design/Algorithm

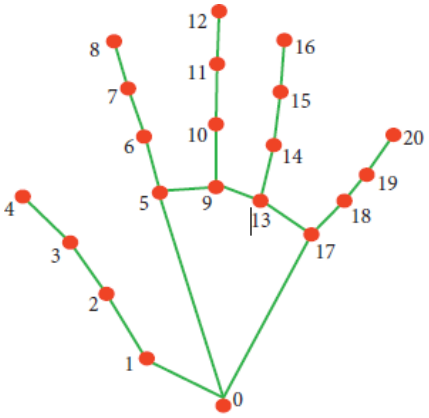
It uses the Media Pipe framework for hand gesture recognition and hand tracking, and the OpenCV library for computer vision. The algorithm uses machine learning concepts to track and recognize hand gestures and hand tips..

MediaPipe is a framework used for applications in machine learning pipelines and is an open source framework from Google. The MediaPipe framework is useful for cross-platform development because the framework is built using time-series data. The MediaPipe framework is multimodal and can be applied to a wide variety of audio and video. The MediaPipe framework was used by developers to create and analyze systems using diagrams, and was also used to develop systems for application purposes. System steps using MediaPipe are executed in a pipeline configuration. The pipelines created can run on different platforms, enabling scalability across mobile and desktop devices. The MediaPipe framework is based on three basic parts. They are a collection of reusable components called Performance Evaluation, Sensor Data Acquisition, and Calculators. A pipeline is a graph made up of components called processors, each connected by a stream through which data packets flow.

finished. Developers can substitute or define custom calculators anywhere in the chart to create their own applications. Combining a calculator and a stream creates a data flow her diagram. Graphs are created using MediaPipe. Each node is a calculator and the nodes are connected by streams.

A single-shot detector model is used to detect and detect hands and palms in real time using A single-shot detector model is used in MediaPipe. First, the hand recognition engine is first trained with a palm recognition model because it is easier to train the palm. Additionally, submaximal suppression is very effective for small objects such as palms and fists. A hand marking model consists of identifying the joint or knuckle coordinates of the hand region. OpenCV is a computer vision library containing image processing algorithms for object detection. OpenCV is a Python programming language library that allows you to develop real-time computer vision applications using computer vision libraries. The OpenCV library is used for image and video processing and analytics such as face detection and object detection.

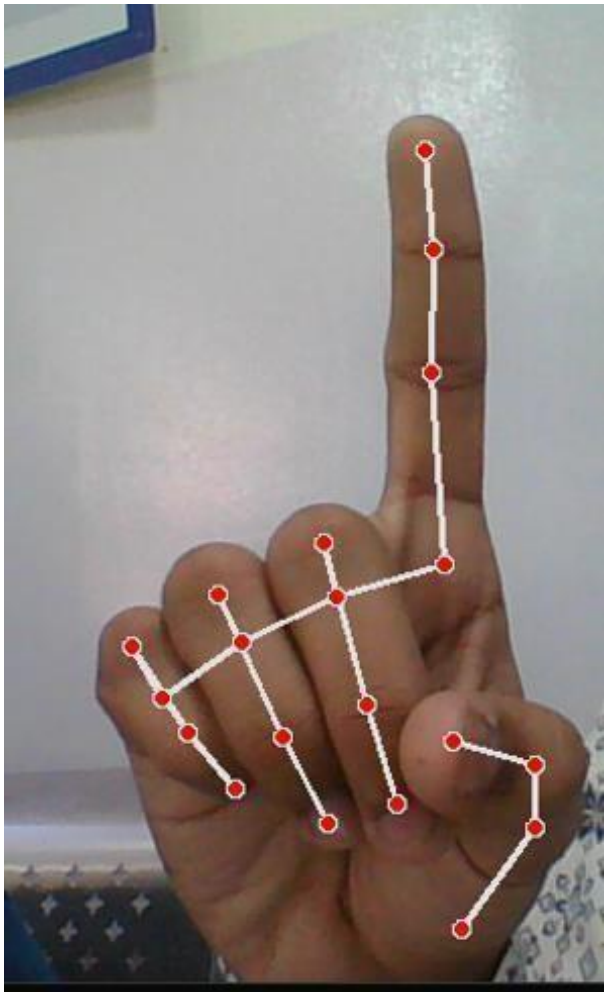




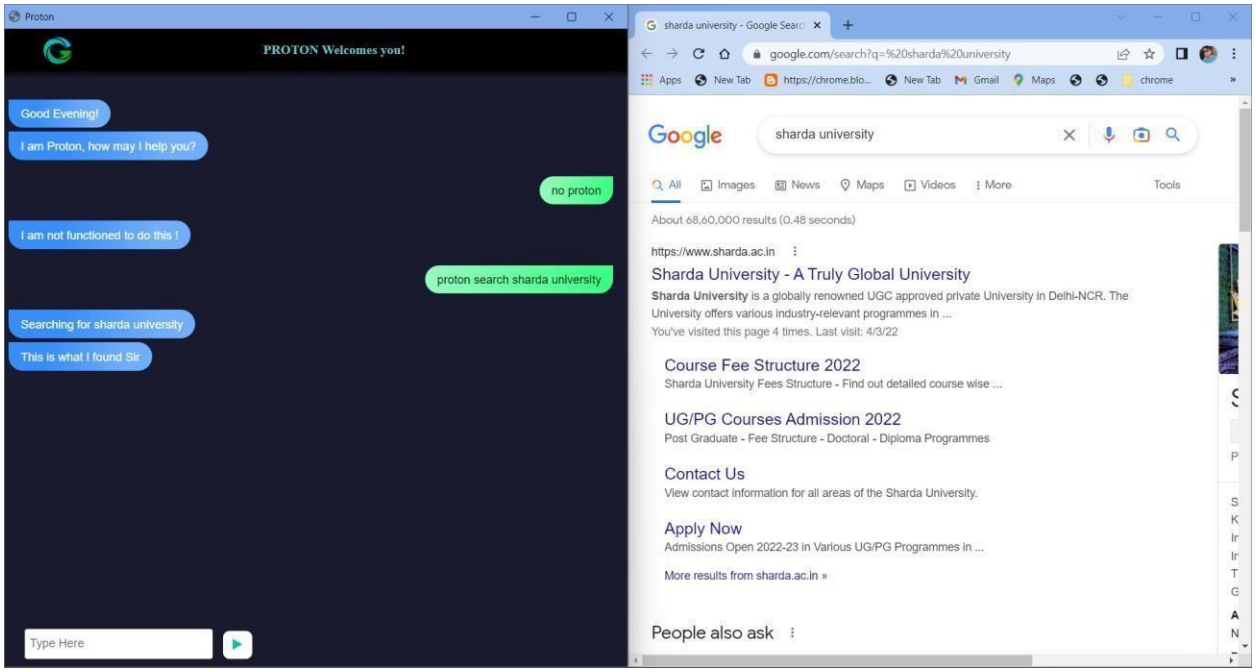
0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER_PIP
11. MIDDLE_FINGER_DIP
12. MIDDLE_FINGER_TIP
13. RING_FINGER_MCP
14. RING_FINGER_PIP
15. RING_FINGER_DIP
16. RING_FINGER_TIP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

Result:

Hand Detected By Webcam



Connected Hand With Mouse



Connected with voice Assistance

Test Cases:

Test case id	Scenario	Boundary value	Expected Result	Actual Result	Status
1.	Used in normal environment.	>90%	In normal environment hand gestures can be recognized easily.	Hand gestures got easily recognized and work properly.	**Passed**
2.	Used in bright environment.	>60%	In brighter environment, software should work fine as it easily detects the hand movements but in a more brighter conditions it may not detect the hand gestures as expected	In bright conditions the software works very well.	**Passed**

3.	Used in dark environment	>30%	In dark environment, It should work properly.	In dark environment software didn't work properly in detecting hand gestures.	**Failed**
4.	Used at a near distance (15cm) from the web cam.	>80%	At this distance, this software should perform perfectly.	It works fine and all features works properly	**Passed**
5.	Used at a far distance(35cm) from the web cam.	>95%	At this distance, this software should work fine.	At this distance, it is working properly.	**Passed**
6.	Used at a farther distance(60cm) from the web cam.	>60%	At this distance, their will be some problem in detecting hand gestures but it should work fine.	At this distance, The functions of this software works properly.	**Passed**

Conclusion:

In this project, I'm working on a system for controlling a mouse pointer with a real-time camera. The system is based on computer vision algorithms and can handle all mouse tasks. However, people's flashes and skin tones are different, so it's difficult to get consistent results. This system makes presentations easier and saves working space. It provides functions such as zooming in and out of the window, closing the window, etc. palm and all fingers. This software was developed to enable people with disabilities to use their desktops and laptops as intelligently as normal people and to consider using new edge technologies. Python
3.7 (64-bit) and open source modules are used to develop this project, making it suitable for future updates.

Future Scope and further enhancement of the Project:

Our goal is to implement additional gestures that will allow users to do more in the future. Effectively. This project proposes the following system. Use the appropriate hand to perform the gesture. So do it to see the possible future more clearly. Online and distance learning, and TV control. Sign language translation is the most efficient application for deaf people. A project must provide a service. purpose for which they were developed. We demonstrate that currently implemented implementation improvement techniques have been successful, allowing both hands to be used for different gestures. Many applications Significant improvement with high speed. Development of hand gesture recognition system. Especially doctors and surgeons are all patients. It is automated with the language assistant function. Actions such as finding a place on Google Maps. Document navigation, start and stop gesture recognition, Google search, sleep/wake voice assistant. This feature saves end users time and effort.

Advantages of this Project:

A) Smart Movement: the current detection process is limited to a 25 cm radius, so it is automatically adjusts the focus ratio to the distance between the user and the webcam, and adaptively zooms in to improve the travel distance. out functionality is required.

B) Better Accuracy and Performance: Response time is highly dependent on device hardware. B. Processor speed, amount of available RAM, and available webcam capabilities. So it's possible that the program will work better if you run it on a decent computer with a webcam that works better in different types of lighting.

C) Mobile Application: In the future, this web application will also be available on his Android device, replacing the touch screen concept with hand gestures.

References

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