

Virtual control mouse using gestures and voice assistant

V. Pethuru¹, V. Hemasai², V. Tejesh³, M. Abhishek⁴,
Mr. K B Rajeswara Rao⁵, Dr. K. Rajendra Prasad⁶

^{1,2,3,4}Students of Department of CSE, NSRIT, Vishakhapatnam, India

^{5,6}Assistant Professor, Department of CSE, NSRIT, Vishakhapatnam, India

Abstract:

We usually interact with computers using a mouse and a keyboard because they need us to touch them. These devices do not work well in places where we cannot touch things or where people need special help. This system is different because it lets us control things without needing to touch any hardware. It uses our voice and the movements of our hands. When we give voice commands it starts looking for hand movements. A webcam watches what our hands are doing.

Keywords: Human Computer Interaction, Virtual Mouse Hand Gesture Recognition, Voice Control, Computer Vision, Speech Recognition.

1. INTRODUCTION

A computer mouse is something we use to control the cursor on a screen and work with the Graphical User Interface. It can tell when we move it around on a surface and it makes the pointer.

There are kinds of computer mice now. The old kind had a rubber ball inside that helped it figure out how we were moving it. Then they made a kind that used a special light to see how we were moving it and that worked better. In the year 2004 they came out with a laser mouse that was more precise and could work well on shiny surfaces.

2.1 Problem Statement

I am not surprised that all technological devices have their limitations especially when it comes to computer devices. After I reviewed types of physical mouse I found some are pretty common. The problem with computer mice is that they can cause discomfort and even injury to users who spend a lot of time using them. Traditional computer mice, carpal tunnel syndrome and other repetitive strain injuries.

The following describes the problem that the current physical mouse suffers from:

1. Computer mouse is subjected to mechanical wear and tear.
2. Computer mouse requires hardware and a surface to work properly.
3. Computer mouse does not adapt easily to environments and the performance of the computer mouse varies depending on the environment.

2.2 System Design

1. System design is about creating a plan for a system that meets needs. It helps businesses or organizations work efficiently.
2. To start you need to figure out what your business or organization requires.
3. Then you can build a system that meets those needs and makes your customers happy.

Good system design means thinking about everything from hardware and software to data storage.

1. Now let's talk about gesture recognition.
2. It has two phases: calibration and recognition.
3. The calibration phase is important.

2.3 Calibration Phase

Real Time Image Acquisition

The program will start of by capturing real-time images via a webcam where it will wait for users' input. The size of the acquired image will be compressed to a reasonable size to reduce the processing loads of processing the pixels within the captured frame.

User Input Acquisition

The program acquires the frames that consist of input submitted by the users, the captured frame will be sent for process where it will undergo a series of transition and calculation to acquire the local landmark points values.

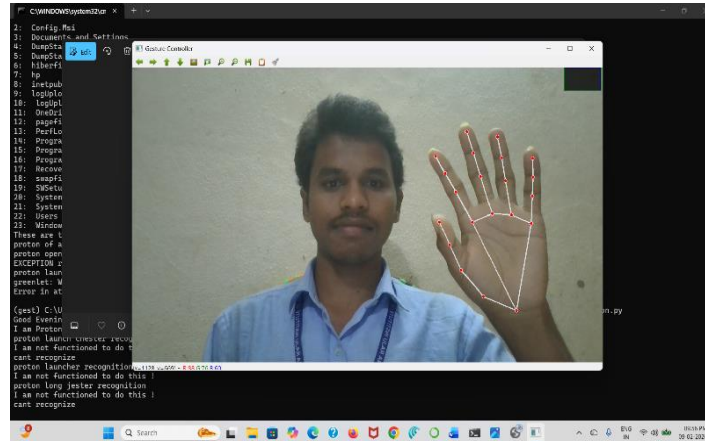


Figure 2.3: Real time acquisition of input

Frame Noise Filtering

Every captured frame consists of noises that will affect the performance and the accuracy of the program; therefore, the frame requires to be noise free. To do that, filters need to be applied on the captured frames to cancel out the unwanted noise.

2.4 Recognition phase

a) Webcam & Variables Initialization

On the early stage of the recognition phase, the program will initialize the required variables which will be used to hold different types of frames and values where each are will be used to carry out certain task.

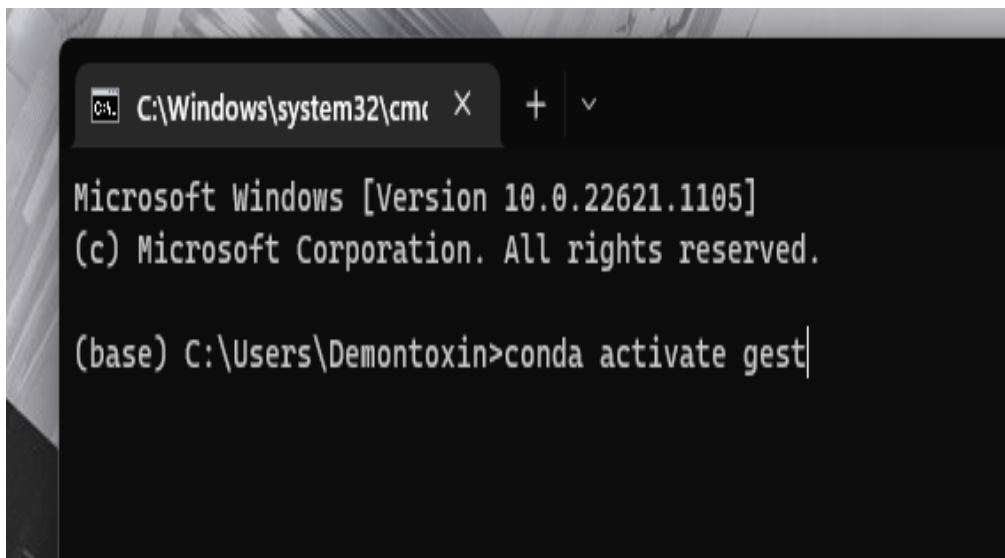
b) Real Time Image Acquisition

The real time image is captured by using the webcam by using `(cv::VideoCapture cap(0);)`, where every image captured are stored into a frame variable `(cv::Mat)`, which will be flipped and compressed to a reasonable size to reduce process load

3. APPLICATION LAYOUT

The application will start of by displaying a console command window; users are required to activate the virtual environment using "conda activate gest" command. "conda activate gest" is a command that activates the "gest" environment in the Conda package manager. Assuming you have already installed Conda, you can activate the "gest" environment. This command will activate the "gest" environment and set it as your current environment. Once activated, you can run any software or programs that are installed in the "gest" environment. To deactivate the environment and return to your base environment, you can simply type: "conda deactivate".

In general, an environment in Conda is a self-contained directory that contains a specific collection of packages and their dependencies. Conda allows you to create multiple environments that can have different versions of packages and dependencies installed, which can be useful for managing different projects or applications that require different software dependencies.

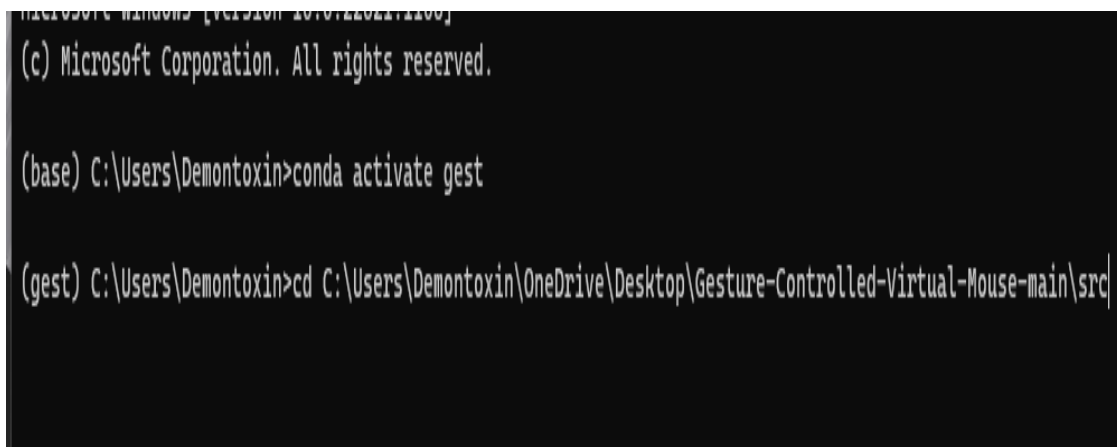


```
C:\Windows\system32\cmd X + v
Microsoft Windows [Version 10.0.22621.1105]
(c) Microsoft Corporation. All rights reserved.

(base) C:\Users\Demontoxin>conda activate gest|
```

Figure 3.0: Activating Environment

The next step is to shift directories to the source folder in a terminal. To do this, type cd followed by the path.



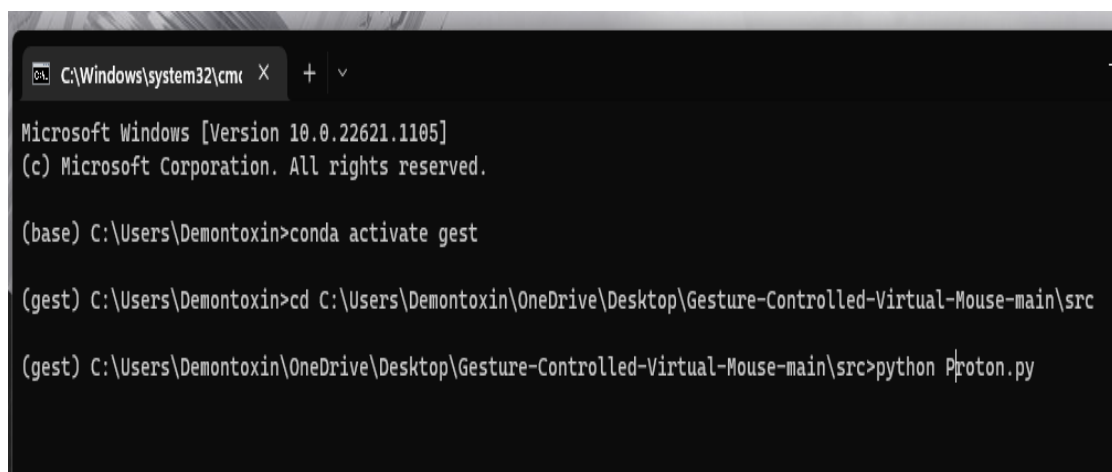
```
Microsoft Windows [Version 10.0.22621.1105]
(c) Microsoft Corporation. All rights reserved.

(base) C:\Users\Demontoxin>conda activate gest

(gest) C:\Users\Demontoxin>cd C:\Users\Demontoxin\OneDrive\Desktop\Gesture-Controlled-Virtual-Mouse-main\src|
```

Figure 3.1 Changing Directory

As this is beginning stage of project we don't have much GUI instead we are using a command line interface to open the Gesture controlled mouse. After typing the path of project file , we have to open the python file,extension “.py”



```
C:\Windows\system32\cmd X + v
Microsoft Windows [Version 10.0.22621.1105]
(c) Microsoft Corporation. All rights reserved.

(base) C:\Users\Demontoxin>conda activate gest

(gest) C:\Users\Demontoxin>cd C:\Users\Demontoxin\OneDrive\Desktop\Gesture-Controlled-Virtual-Mouse-main\src

(gest) C:\Users\Demontoxin\OneDrive\Desktop\Gesture-Controlled-Virtual-Mouse-main\src>python Proton.py
```

Figure 3.2: Opening Python Files.

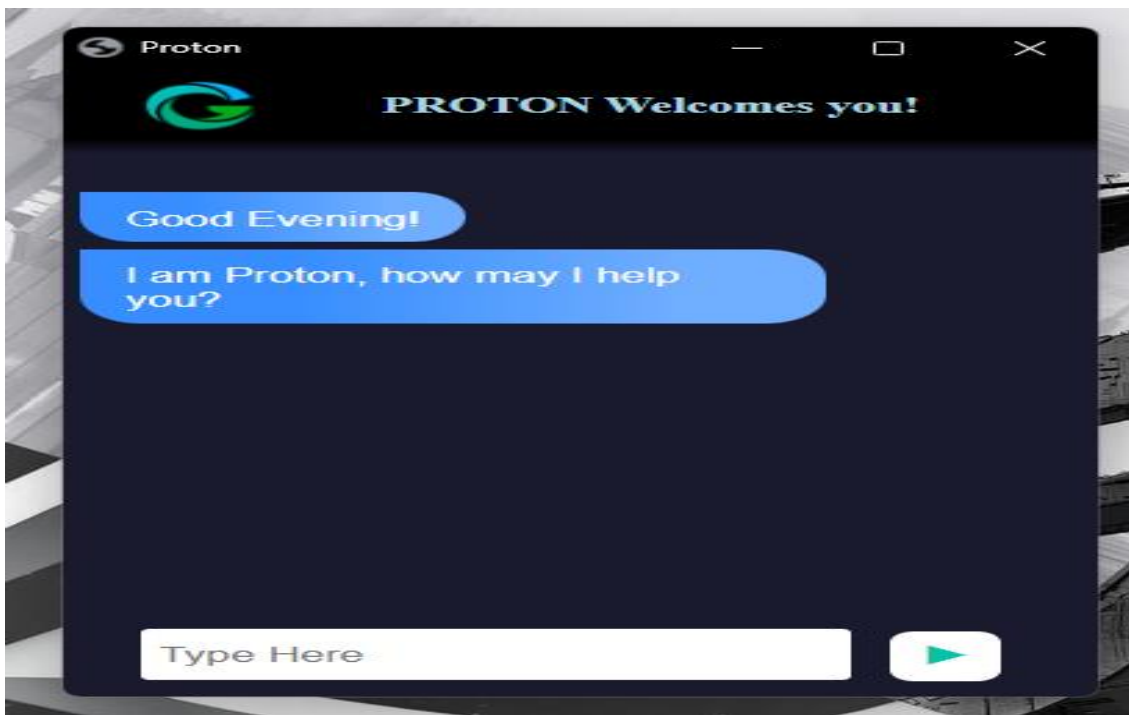


Figure 3.3 Proton voice assistant.

4.1 METHODOLOGY

For this project we'll be using the Agile Software Development methodology approach in developing the application. The stated approach is an alternative to the traditional waterfall model that helps the project team respond to unpredictability through incremental and iterative work. It promotes adaptive planning, evolutionary development, early delivery, continuous improvement, and encourages rapid and flexible respond to change. The following describes the principles of the Agile Software Development methodology:

- a. Satisfy the customer by early and continuous delivery of workable software.
- b. Encourage changes of requirement.
- c. Workable software is delivered frequently.
- d. Continuous collaboration between the stakeholders and the developers.
- e. Project are developed around motivated individuals.
- f. Encourage informal meetings.
- g. Operational software is the primary indicator of development.
- h. Sustainable development, able to maintain a constant pace.
- i. Continuous attention to technical excellence and good design
- j. Simplicity
- k. Self-organizing teams

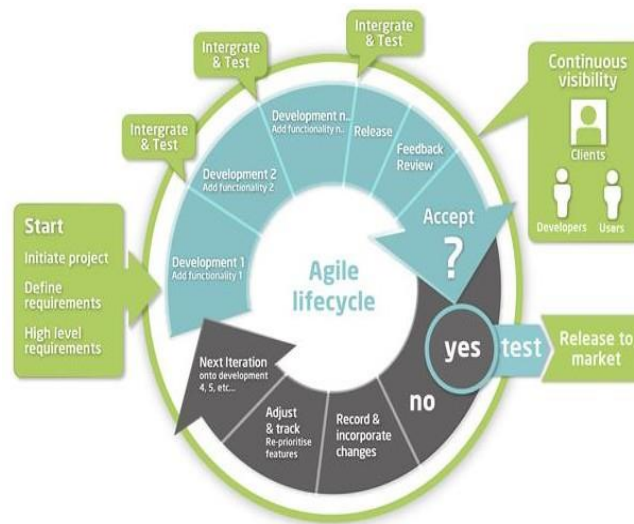


Figure 4.1: Agile method overview

4.2 Software Requirement

Software requirements deal with defining software resource requirements and prerequisites that need to be installed on a computer to provide optimal functioning of an application. These requirements or prerequisites are generally not included in the software installation package and need to be installed separately before the software is installed.

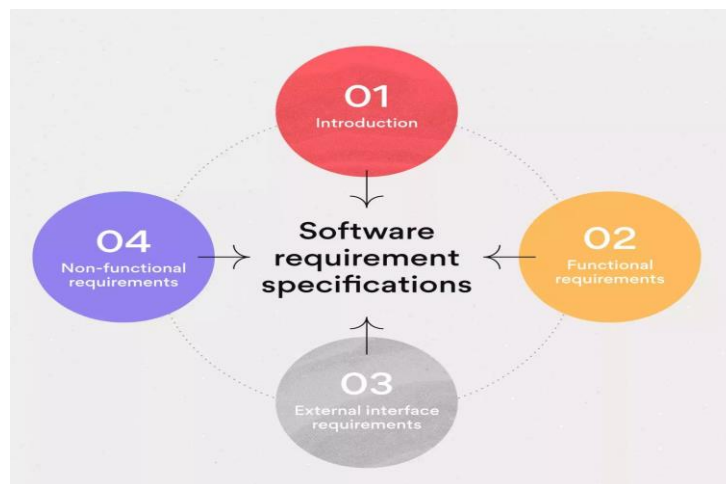


Figure 4.2: Software Requirement Specification.

4.2.1 MediaPipe:

MediaPipe is a Framework for building machine learning pipelines for processing time-series data like video, audio, etc. This cross-platform Framework works in Desktop/Server, Android, iOS, and embedded devices like Raspberry Pi and Jetson Nano. The MediaPipe perception pipeline is called a Graph. Let us take the example of the first solution, Hands. We feed a stream of images as input which comes out with hand landmarks rendered on the images.

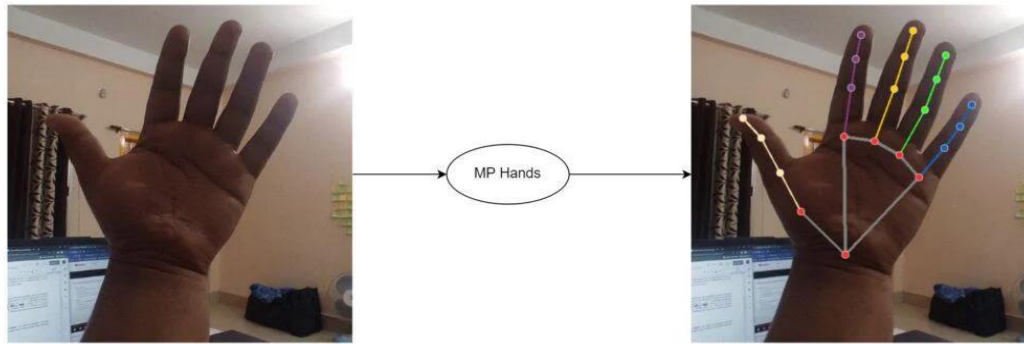


Figure 4.2.1: MediaPipe Hands.

In this example, we use Mediapipe's pre-built hand tracking model to detect hand landmarks in real-time video stream captured from a webcam. We then draw circles at the landmarks' positions to visualize their location. Overall, Media pipe is a powerful and versatile framework for building computer vision applications, and its ease of use and flexibility make it an excellent choice for developers of all skill levels

4.2.2 Open CV Library

OpenCV are also included in the making of this program. OpenCV (Open Source Computer Vision Library) is a library of programming functions for real time computer vision. OpenCV have the utility that can read image pixels value, it also has the ability to create real time eye tracking and blink detection. OpenCV is a huge open-source library for computer vision, machine learning, and image processing. OpenCV supports a wide variety of programming languages like Python, C++, Java, etc. It can process images and videos to identify objects, faces, or even the handwriting of a human.

5.SPEECHRECOGNITION:

Speech Recognition is an important feature in several applications used such as home automation, artificial intelligence, etc. This article aims to introduce how to make use of the SpeechRecognition library of Python. This is useful as it can be used on microcontrollers such as Raspberri Pi's with the help of an external microphone.

Speech recognition is the process of converting spoken words into written text. Python has a built-in module for speech recognition called the Speech Recognition module.

Features:

- It is a cross-platform, open-source library for speech recognition and natural language processing
- Supports multiple languages such as English, Spanish, French, German, and more.
- Supports multiple platforms such as Windows, macOS, Linux, and more.
- Supports multiple formats such as WAV, AIFF, MP3, and more.

Installation:

Install the Speech Recognition module by downloading it from PyPI or by using pip. the following must be installed: Python Speech Recognition module:

pip install SpeechRecognition

Uses:

- Import the Speech Recognition module into your project.
- Create a Recognizer object and set some parameters

5.1 ALGORITHM

An algorithm is a step-by-step procedure or a set of instructions that helps solve a specific problem or complete a task. In computer science, algorithms guide computers on how to perform a task or process data. Typically, an algorithm includes a series of clear steps that are carried out in a certain order to yield a desired outcome. Algorithms can be represented in different formats, including pseudocode, flowcharts, or programming languages. The steps in an algorithm must be clear and straightforward, so anyone following them can achieve the same result.

5.1.2 Convolution Neural Network

A Convolutional Neural Network (CNN) is a type of Deep Learning neural network architecture commonly used in Computer Vision. Computer vision is a field of Artificial Intelligence that enables a computer to understand and interpret the image or visual data.

5.1.3 Viola Jones Algorithm

Viola Jones algorithm is named after two computer vision researchers who proposed the method in 2001, Paul Viola and Michael Jones in their paper, "Rapid Object Detection using a Boosted Cascade of Simple Features". Despite being an outdated framework, Viola-Jones is quite powerful, and its application has proven to be exceptionally notable in real-time face detection. This algorithm is painfully slow to train but can detect faces in real-time with impressive speed.

6. IMPLEMENTATION AND TESTING;

A neutral gesture in a virtual mouse is a position or movement that does not trigger any specific action or command. In other words, it is a gesture that indicates the user's intention to neither click, drag, nor perform any other action with the virtual mouse. In most virtual mouse software, the neutral gesture is typically represented by placing the cursor .

6.1 Cursor Movement:

When you use a mouse you can move the cursor around by clicking and dragging it with the virtual mouse. You can also use things like a touch pad or trackball to move the cursor.

6.2 Cursor Movement of Virtual Mouse

Moving the cursor is when you use your mouse to make the cursor on your computer go where you want it to. The cursor can go up down side to side or else you want and it can go as fast or as slow as you like. You can use the cursor to look at all your menus and pick things on your computer.

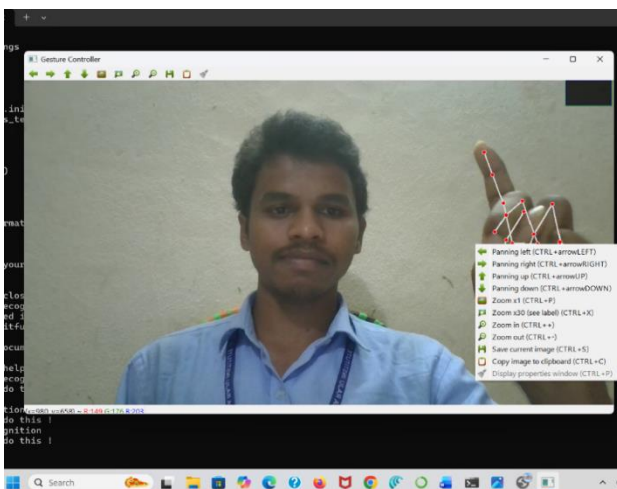


Figure 5.5 : Left Click Gesture



Figure 5.6: Gesture to increase and decrease sound

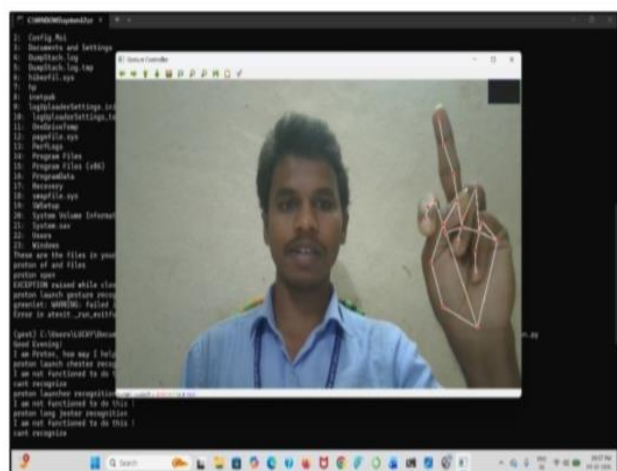


Figure 6.3: Right Click Gesture

6.4 Voice Commands

Some computer programs can figure out what you are saying. This means you can move the cursor and do things with your voice. You can tell the computer what to do. It will do it.

Virtual mouse outputs are really helpful. They make it easy to use a computer with your voice or other devices. You can use your voice to control the computer. Voice commands are an example of mouse outputs. Voice commands are really useful. They make things easy, for you.

Working of Voice Command System

1. The user gives a voice command through a microphone.
2. The system converts speech into text using speech recognition.
3. The command is matched with predefined instructions.
4. The corresponding action is executed.
5. The result is displayed on the website interface

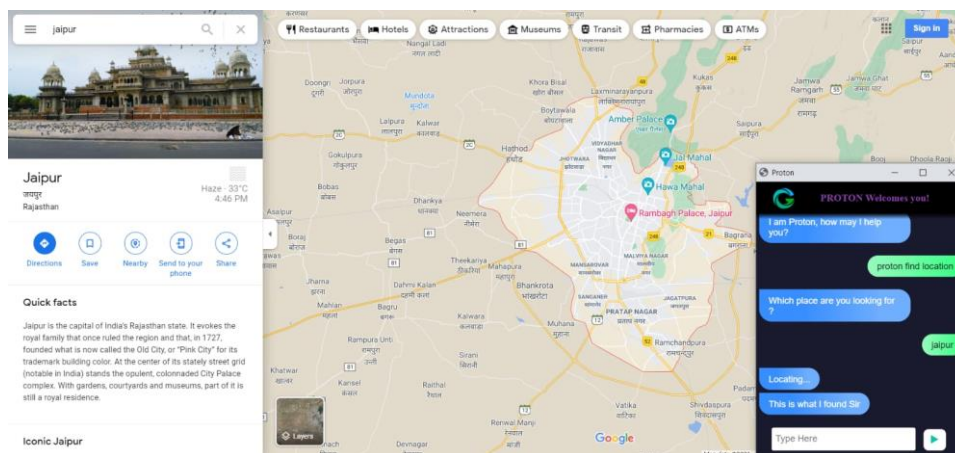


Figure 6.4.1: Voice assistant locating a place using voice command

7. CONCLUSION

The physical mouse will probably be replaced by a mouse in Human-Computer Interactions. This is because we can use our fingers to control the mouse at any time. A program was created to recognize gestures and turn them into mouse actions. This program is meant to replace the mouse without making it less accurate or efficient. To make this program work well a few strategies have to be used. Precision and efficiency are very important when it comes to making the program as helpful as a mouse. The main goal of this technique is to reduce and stabilize the movement of the cursor. This is because even small movements can cause the cursor to move when we do not want it to.

First the coordinates of the fingertips that control the cursor are averaged. This is done by taking a collection of coordinates and finding the average. A number of gesture combinations were also created. These combinations include distance calculations between two fingers. This is because the distance between the fingers affects the mouse function. The goal of this implementation is to make the program easier to use. We want to make it so that people do not have to try a lot of things to get the mouse to work.

New technologies have improved our lives. Made us more productive. Therefore, we should not be afraid to accept changes and let go of old technology. We should be open to ideas and changes so that we can live more effective and productive lives. The AI virtual mouse system is meant to replace the mouse with hand gestures. This system can be used with a webcam or a camera that is built into the computer. The camera recognizes hand movements. Processes them to perform certain mouse actions. From the results of the model we can see that the AI virtual mouse system works well. It is more accurate than models and gets rid of many of their problems. The AI virtual mouse can be used in real-life situations because it is more accurate.