भारतीय प्रौद्योगिकी संस्थान
काशी हिन्दू विश्वविद्यालय

IIT INDIAN INSTITUTE OF TECHNOLOGY
BANARAS HINDU UNIVERSITY

Exploratory Project Report

SEMESTER IV 2021-22

VOLUME CONTROL USING HAND GESTURE AND FINGER COUNTER IN HAND

Project done under the supervision of Dr. Satyabrata Jit by:

Avinash Singh (20095020) **Korivi Vedharshini (20095055)** **Nipun Khanduja (20095073)**

Introduction

The ability to perceive the shape and motion of hands can be a vital component in improving the user experience across a variety of technological domains and platforms. While coming naturally to people, robust real-time hand perception is a decidedly challenging computer vision task, as hands often occlude themselves or each other (e.g., finger/palm occlusions and handshakes) and lack high contrast patterns.

As we know, the vision-based technology of hand gesture recognition is an important part of human-computer interaction (HCI). In the last decades, keyboard and mouse play a significant role in human-computer interaction.

However, owing to the rapid development of hardware and software, new types of HCI methods have been required. In the faster growing world of automation technologies such as gesture recognition receive great attention in the field of HCI and it also seeks great demand as well.

Aim and Objectives

The project aims to create interaction between machine hardware and hand gestures using landmarks on hand to minimize the use of hardware. Our objective is to create some applications of this through which we can illustrate the use of our project.

This project includes:

- Volume controller using hand movement.
- Counter using hand landmarks of fingers.

Basic Workflow

Firstly, the program detects landmarks on the hand, there are in total 21 of these land marks as we will see in the further slides. Then, the positions of these landmarks are stored in the program. After that, a hand module is created which gives the positions of these landmarks on the hand in the live camera feed. We will use this hand module to do two tasks- Controlling volume and counting numbers.

Libraries Used

OpenCV: It is the huge open-source library for computer vision, machine learning, and image processing. It plays a major role in real-time operation which is very important in today's systems. By using it, one can process images and videos to identify objects, faces, or even handwriting of a human. When it is integrated with various libraries, such as NumPy, python is capable of processing the OpenCV array structure for analysis. To Identify image pattern and its various features we use vector space and perform mathematical operations on these features.

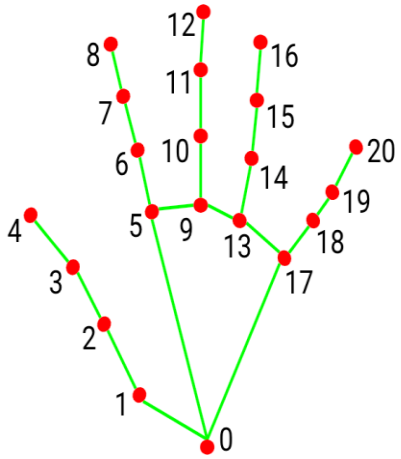
Media Pipe: Media Pipe Hands is a high-fidelity hand and finger tracking solution. It employs machine learning (ML) to infer 21 3D landmarks of a hand from just a single frame. When current state-of-the-art approaches rely primarily on powerful desktop environments for inference, our method achieves real-time performance on a mobile phone, and even scales to multiple hands.

Models Implemented

Palm Detection Model: To detect initial hand locations, we designed a single-shot detector model optimized for mobile real-time uses in a manner similar to the face detection model in Media Pipe Face Mesh. Detecting hands is a decidedly complex task: our model has to work across a variety of hand sizes with a large scale span (~20x) relative to the image frame and be able to detect occluded and self-occluded hands. The lack of high contras features in hands makes it comparatively difficult to detect them reliably from their visual features alone. Instead, providing additional context, like arm, body, or person features, aids accurate hand localization.
Palms can be modelled using square bounding boxes (anchors in ML terminology) ignoring other aspect ratios, and therefore reducing the number of anchors by a factor of 3-5. Second, an encoder-decoder feature extractor is used for bigger scene context awareness even for small objects. Lastly, we minimize the focal loss during training to support a large number of anchors resulting from the high scale variance.

Hand Landmark Model: After the palm detection over the whole image our subsequent hand landmark model performs precise key point localization of 21 3D hand-knuckle coordinates inside the detected hand regions via regression, that is direct coordinate prediction. The model learns a consistent internal hand pose representation and is robust even to partially visible hands and self-occlusions.

To obtain ground truth data, we have manually annotated around 30000 real-world images with 21 3D coordinates, as shown below (we take Z-value from image depth map, if it exists per corresponding coordinate). To better cover the possible hand poses and provide additional supervision on the nature of hand geometry, we also render a high-quality synthetic hand model over various backgrounds and map it to the corresponding 3D coordinates.



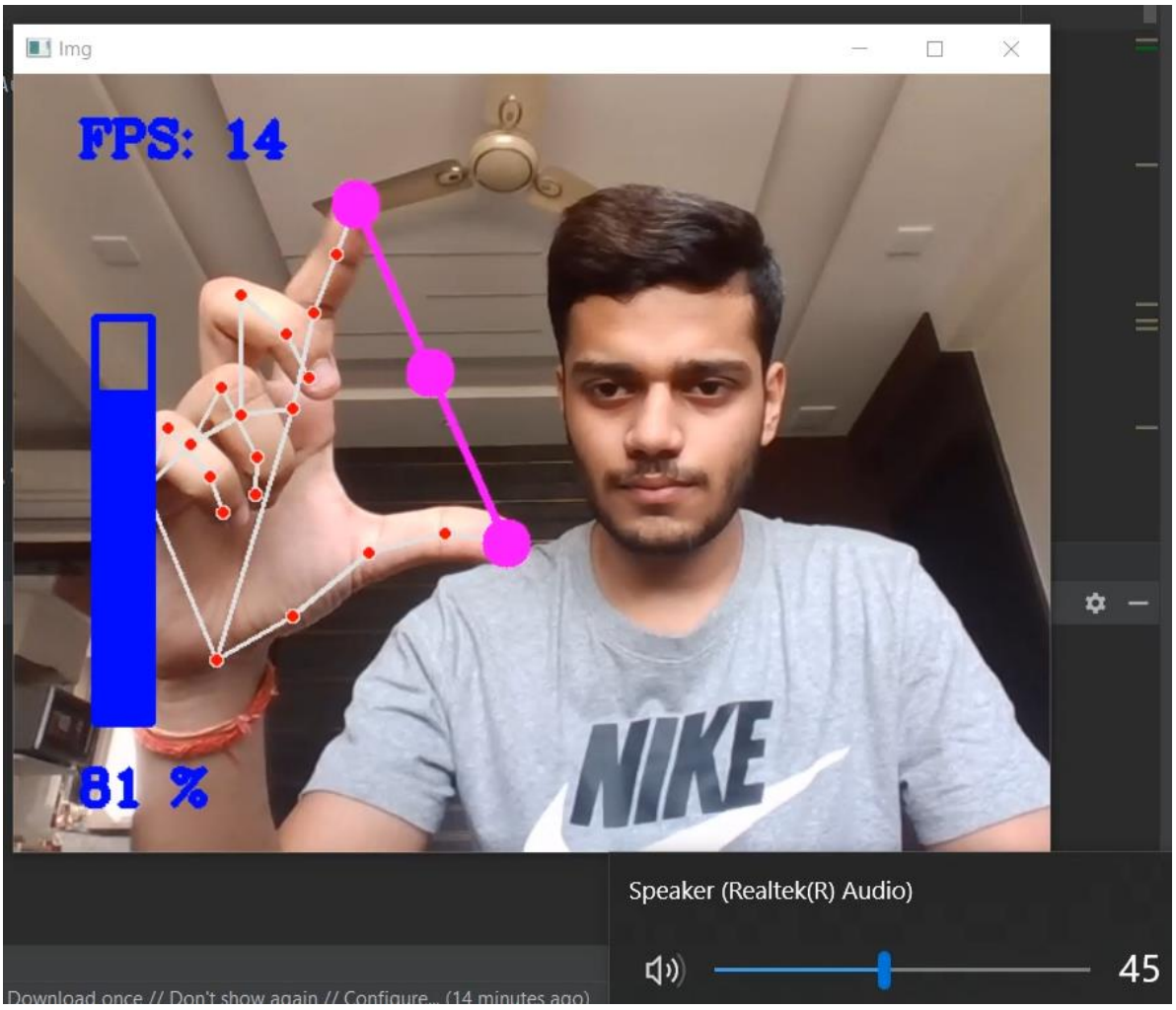
0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER_PIP
11. MIDDLE_FINGER_DIP
12. MIDDLE_FINGER_TIP
13. RING_FINGER_MCP
14. RING_FINGER_PIP
15. RING_FINGER_DIP
16. RING_FINGER_TIP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

RESULT: Volume Control using Hand Movement

When we run the code, we can see the feed.

Taking a hand in front of camera shows us 21 landmarks of the palm which we are shown as red dots which are connected by white lines.
The pink line between thumb and index finger shows the distance between those two.

As we can see from the below gif moving finger up and down (i.e Changing distance) help us to adjust the volume of our laptop.



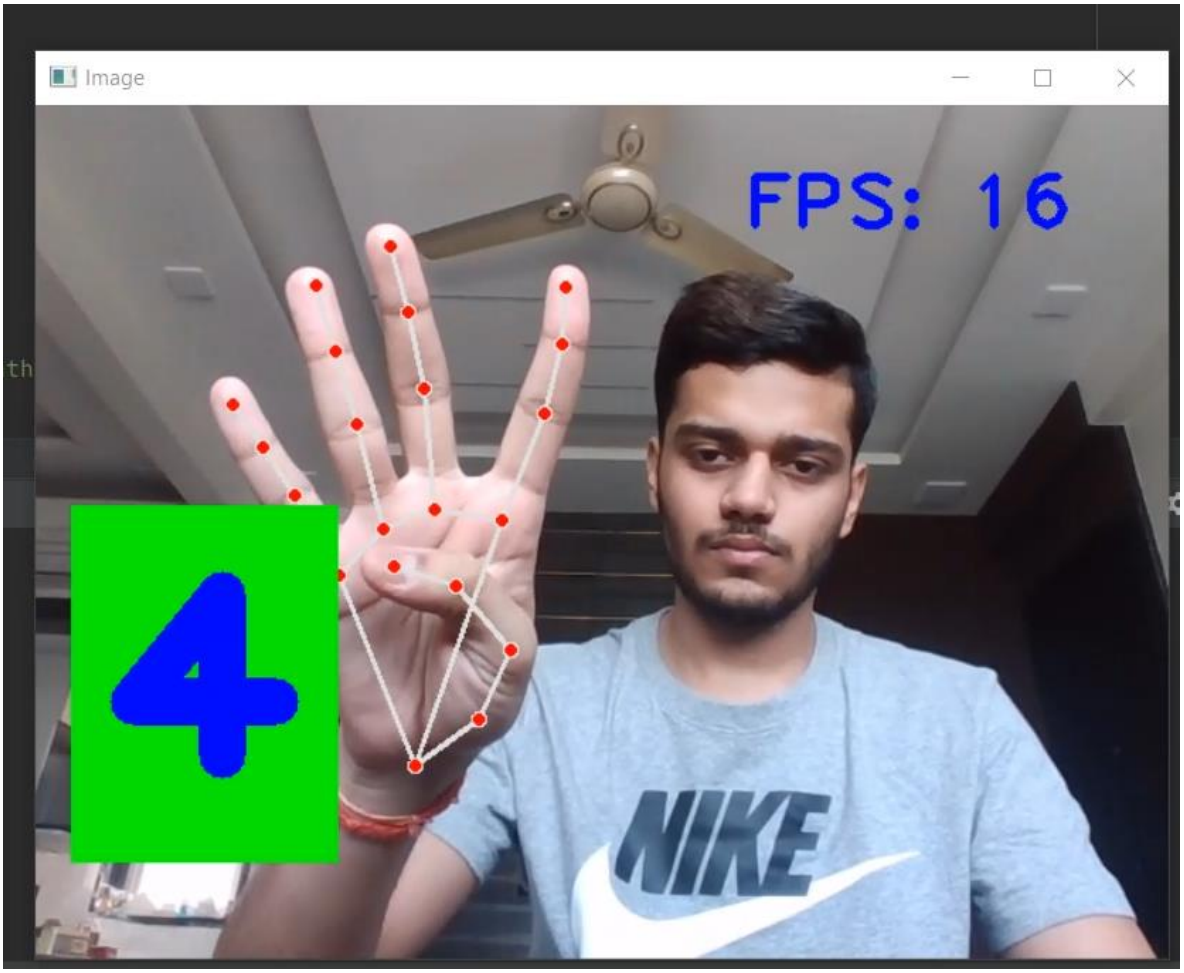
This thing will help us to reduce dependency on hardware to do our daily task and make life simpler.

RESULT: Counter Using Hand Landmarks on Finger

When we run the code, we get to see the live feed.

Taking a hand in front of the camera shows us 21 landmarks of the palm which we are shown as red dots which are connected by white lines.

From below GIF we can observe that as if ID of the top of finger
Goes above the other ID's of the same finger that present below it then it considered as unfolded and is counted.
And when it goes below that it is considered as folded.



This is just an example of what we can do by changing the relative positions of 21 points. We can use the positions to give different gestures of hand as input. These gestures can be fed as a command to certain devices through which can control hardware simply by changing our hand gesture.

Conclusion and Future Insights

- We can conclude that our exploration of this project can be used in various fields in daily life. This will reduce the human dependency on hardware.

- In this time of pandemic, we are avoiding touch to any surface. By our project we can make smart control systems which minimizes human touch.

- This has tremendous future use like - smart homes, high security systems etc.

- In the medical fields Hand Gesture may also be experienced in terms of Robotic Nurse and medical assistance. As the Technology is always revolving and changing the future is quite unpredictable but we have to be certain the future of Gesture Recognition is here to stay with more and eventful and Life touching experiences.

- This is moving at tremendous speed for futuristic products and services and major companies are developing technology based on the hand gesture system and that includes companies like Microsoft, Samsung, Sony.

- It's a brilliant feature turning data into features with a mix of technology and Human wave. This also reduces our dependency on hardware as well.

- Smart phones have been experiencing an enormous amount of Gesture Recognition Technology with look and views and working to manage the Smartphone in reading, viewing and that includes what we call touch less gestures.