

VIRTUAL MOUSE USING HAND GESTURE RECOGNITION

Renuka TR¹, Samvrutha Shaji², Sreelakshmi E³

¹ Student, CSE, Government Engineering College Sreekrishnapuram, Kerala, India

² Student, CSE, Government Engineering College Sreekrishnapuram, Kerala, India

³ Student, CSE, Government Engineering College Sreekrishnapuram, Kerala, India

ABSTRACT

The purpose of this project is to develop a virtual mouse that can be controlled using hand gestures. The virtual mouse will be able to replicate the functions of a physical mouse, such as moving the cursor, clicking, and scrolling. The system will be implemented using a camera to capture hand gestures and machine learning algorithms to interpret and translate them into mouse actions. The project aims to provide a convenient and intuitive way for users to interact with their computers, particularly for those who may have difficulty using a traditional mouse. To achieve the desired functionality, the system will use a combination of computer vision techniques and machine learning algorithms. The project will be evaluated through user testing, in which a group of participants will be asked to perform a series of tasks using the virtual mouse. The results of the user testing is used to assess the accuracy and usability of the system, and to identify any areas for improvement. Overall, the virtual mouse using hand gestures has the potential to provide a convenient and intuitive way for users to interact with their computers, and could be particularly beneficial for those with physical disabilities.

Keyword: - *Virutal Mouse, OpenCV, Mediapipe, PyaotoGUI, Hand detection*

1. Introduction

In recent years, the applications such as human-computer interface (HCI) and robot vision have been activated in research areas. As in virtual environment (VE) applications, they offer the opportunity to integrate various latest technologies to provide a more immersive user experience. A virtual mouse with HCI arises much attention and it could substitute the function of traditional mouse and touch panel. To realize this device, the techniques based on computer vision are involved. A vision-based HCI system with hand gesture analysis is introduced. Our design on this work explores the low cost and high-performance virtual mouse. All the functions on computer mouse are considered. We capture the hand gesture by a single web camera for natural and intuitive human-computer interaction. With this common environment, the system can provide high recognition result even in some harsh background environment.

2. Proposed System

In the proposed system, the Object tracking method has been used to track hand gestures in order to perform corresponding mouse action. By using the Object tracking system, the mouse and its basic operations like mouse pointing, left and right clicking, and scroll up and down can be controlled. Computer vision techniques are used for gesture recognition. OpenCV consists of a package called video capture which is used to capture data from a live video. System's webcam is used for tracking hand gestures. The proposed system will detect hand gestures and perform mouse actions without using color caps. Media Pipe is a cross-platform pipeline framework to build custom machine learning solutions for live and streaming media. The framework was open-sourced by Google and is currently in the alpha stage. The MediaPipe framework is mainly used for rapid prototyping of perception pipelines

with AI models for inferencing and other reusable components. It also facilitates the deployment of computer vision applications into demos and applications on different hardware platforms. In the proposed system, we use the Python language for the implementation. We use the OpenCV library for image processing and pyautoGUI library for mouse control. Implementation of the system includes two methods, one for hand tracking and another to perform corresponding mouse action.

2.1 Hand Detection

Once the webcam is opened, the user can show his/her hand in front of the webcam and the system will detect their hand. Then user can move their hand and can perform different actions. User's hand is detected using the Mediapipe library in python. This library contains two modules Palm detection module and Hand landmark detection module.

Initially we train a palm detector instead of a hand detector, since estimating bounding boxes of rigid objects like palms and fists is significantly simpler than detecting hands with articulated fingers. In addition, as palms are smaller objects, the non-maximum suppression algorithm works well even for two-hand self-occlusion cases, like handshakes. Moreover, 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 (similar to the RetinaNet approach). Lastly, we minimize the focal loss during training to support a large number of anchors resulting from the high scale variance. With the above techniques, we achieve an average precision of 95.7 in palm detection.

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.

2.2 Different Hand Gestures

Mouse movement is performed by using the index finger. After detecting coordinates of index finger using webcam, the mouse will be moved to respective position on the screen. The relative position is calculated using screen size as well the frame size which is popped when webcam is on.

Various click operations are performed by detecting different hand gestures. Initially the fingertips of five fingers will be detected and then based on the distance between the tips various click operations are performed. Left click operation will be performed on the display if the distance between tip of thumb and tip of index is smaller than the threshold and all the other fingers will be ignored. The right click operation will be performed if the distance between the tip of thumb finger and tip of middle finger is smaller than the threshold. The double click operation will be performed if the distance between the tip of thumb finger and tip of ring finger is smaller than the threshold. The threshold for click operations is generally set to 20.

Scroll up can be performed by keeping the distance between index finger tip and middle finger tip below threshold. Scroll down can be performed by keeping the distance between ring finger tip and middle finger tip below threshold. Threshold is set to 30.

Drag and drop of files can be performed by reducing the distance between thumb finger and pinky finger below a threshold of value 5. If the fingers are too close pyautogui.dragTo() function will drag the corresponding to the location of pinky finger in specified duration.

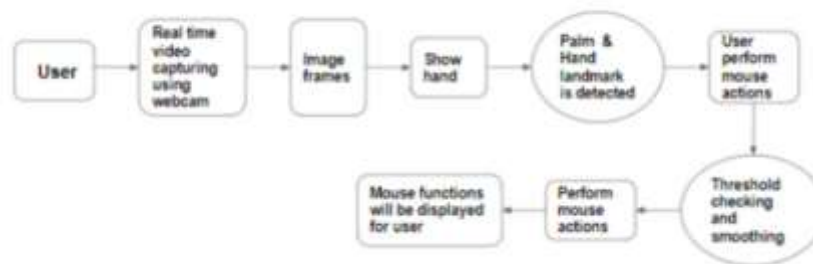


Fig -1: Data flow diagram

3. Methodology

We use mainly three libraries for implementing this system. They are OpenCV, Mediapipe and PyautoGUI. And some basic mathematical calculations are also used to determine the relative positions of our finger with respect to the computer screen.

3.1 Opencv

OpenCV (Open-Source Computer Vision Library) is an open-source computer vision and machine learning software library. The purpose of computer vision is to understand the content of the images. It extracts the description from the pictures, which may be an object, a text description, a three-dimensional model, and so on. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products. Being an Apache 2 licensed product, OpenCV makes it easy for businesses to utilize and modify the code.[1]

3.2 MediaPipe

MediaPipe is a cross-platform pipeline framework to build custom machine learning solutions for live and streaming media. The framework was open-sourced by Google and is currently in the alpha stage. The MediaPipe framework is mainly used for rapid prototyping of perception pipelines with AI models for inferencing and other reusable components. It also facilitates the deployment of computer vision applications into demos and applications on different hardware platforms. The configuration language and evaluation tools enable teams to incrementally improve computer vision pipelines.[3]

3.3 PyautoGUI

PyAutoGUI lets your Python scripts control the mouse and keyboard to automate interactions with other applications. The API is designed to be simple. PyAutoGUI works on Windows, macOS, and Linux, and runs on Python 2 and 3. PyAutoGUI has several features such as moving the mouse and clicking in the windows of other applications, Sending keystrokes to applications (for example, to fill out forms), Take screenshots, and given an image (for example, of a button or checkbox), and find it on the screen, Locate an application's window, and move, resize, maximize, minimize, or close it (Windows-only, currently), Display alert and message boxes.[2]

4. RESULT

The idea behind making this project was the ease of use for every individual. Replacement of the mouse with any object so that they can use this function for what they desire from anywhere concerning be in the frame of a web camera. The use of this project in real-time is vast. We provide several types of hand gestures to perform click scroll and drag mouse functions in better accuracy and efficiency. In each gesture, we test 100 times for each gesture and measure the accuracy of hand gesture recognition.



Fig -2: Mouse movement



Fig -3: Left click



Fig -4: Right click



Fig -5: Double click



Fig -6: Scroll up



Fig -7: Scroll down



Fig-8: Drag and Drop

5. CONCLUSIONS

In this project, we proposed the system to recognize hand gesture and replace the mouse function, which includes the cursor movement, various click operations, scroll up and scroll down. Gesture recognition has been the best way for any interaction between machines and humans. For developing alternative human- computer interaction modules, gesture recognition is also essential. It has enabled us to interface with different computer systems more naturally. Gesture recognition can be deemed to be useful for many applications for people with disabilities.

6. REFERENCES

- [1]. https://docs.opencv.org/4.x/dd/d43/tutorial_py_video_display.html
- [2]. <https://pypi.org/project/PyAutoGUI/>
- [3]. https://developers.google.com/mediapipe/solutions/vision/hand_landmarker