International Journal of Advanced Scientific Research and Management, Volume 9 Issue 5, June 2024

IJASRM

www.ijasrm.com

ISSN 2455-6378

Real-Time Hand Gesture Controlled Virtual Mouse using Open Source Framework

Rohan Pinto[1], Jennifer C Saldanha[2] and Vijay Ganesh P C[3]

[1 2 3] Department of Electronics & Communication Engineering, St Joseph Engineering College, Mangalore, Karnataka State, India.

Abstract

In this modern world, it is unfeasible to imagine life without computers. To make our work easier we may interact multiple times with computer. For research scholars, human-computer interaction has become a prominent topic. In recent times vision-based gesture recognition technology has gained more heed towards friendly HCI interfaces. Hand gestures are efficient for human-computer interaction. Hand gestures offer higher dimensionality and can also control many applications intuitively. This work is on AI virtual mouse using hand gestures and hand tracking.

Keywords: *Human-Computer Interaction, Hand Tracking, Hand Gesture, Virtual mouse, Frame Extraction.*

1. Introduction

In this technical world, the computer is considered as one of the greatest inventions of mankind. Nowadays individual of every age group uses computer everyday, which has become a crucial part. We may interact multiple times with computer to make our work easier. Thus, human computer interaction has become a well-known topic for researchers [1].

The evolution of computer has made vast growth but still, human computer interaction uses input device such as mouse and keyboard. For friendly HCL interfaces, vision-based gesture recognition technology has gained more attention [2]. For human computer interaction, hand gestures are adapted way and they offer higher dimensions and can also control many applications intuitively. National user interface (NUI) uses hand which has become a trendy topic [3]. In this work, the hand is used as a direct input device. For instance, Smartphone contains a touch screen embedded in it. Here the main intention is to interact without direct contact with the surroundings devices.

The most intuitive interface for selection, hand pointing is the basic gesture. Here the challenging task is the appraisal of three-dimensional hand pointing direction accurately and automatically from video input, because of the great variation and suppleness of hand movement and identical features of joint parts of the hand [4].

In a similar way Open CV, media pipe, autopy and pyautogui are used in the virtual mouse that we are implementing in this work. Open CV is a package that is available in python for computer vision that is used for capturing images through Webcam [5]. An Open CV contains a package called video capture which extracts data from a live video. Pyautogui is for specifying keyboard and mouse functions, it is a Python library [6]. MediaPipe is a prebuilt Python package that offers ready-to-use yet customizable Python solutions. With the help of hand gestures virtual mouse permits the user to control the mouse. For hand gesture recognition, the system webcam is used for tracking hand gesture. Computer vision techniques are used for gesture recognition [7].

2. Review of Literature

The mouse has been a vital tool for human-computer interaction, with wired, wireless, and Bluetooth variations requiring power to connect a dongle to a PC. The proposed work uses the latest technology in computer vision and machine learning to recognize hand movements without needing extra equipment. Compatible with CNN models through mediapipe implementation, it enables various actions via various hand movements, all on a single computer. Utilizing a camera as the input device, the system relies on Python and OpenCV. The camera’s output is displayed on the system’s screen, allowing users to fine-tune their interactions [8]

Elderly people face unique challenges when using conventional computer interfaces. Therefore, there is an essential need to model a system for such people for easy accessing of modern computer technologies. This paper presents an inventive solution, the "Gesture and Voice Controlled Virtual Mouse" designed to improve the digital interaction experience for older individuals facing challenges with traditional computer input peripherals. Employing advanced

https://doi.org/10.36282/IJASRM/9.6.2024.1944

technology, this work establishes an intuitive interface using natural gestures and vocal commands. Gesture recognition employs cutting-edge computer vision and machine learning models, for accurate interpretation of automatic hand gestures extraction which is implemented using MediaPipe framework on top of pybind11 along with OpenCV. Voice recognition integrates automatic speech recognition and natural language processing to understand spoken commands effectively. The speech assistant Proton recognizes the commands through pyttsx3 library of python and perform the action required. The experimentation results demonstrated the efficacy of the proposed model in accurate prediction of hand gestures and voice commands [9].

Over the years, presenting the content on chalk boards or through power point presentations have been following in many sectors. Digitalization of things across the world can be seen these days but having an idea to make model which operates on gestures, that doesn't require any physical contact with keyboard, mouse or any kind of hardware devices drives the attention of many users across the globe. Interaction with computers using gestures can be implemented for text displaying which helps people who are having physical disability of joints can use this model for their presentation. This idea also helps for electronic waste management and provides a healthy environment for future generations. The aim of this research is to brief about the idea to mount digitalization using gestures in educational and IT sectors [10].

Gesture-controlled laptops and computers have recently gained a lot of attraction. Leap motion is the name for this technique. Waving our hand in front of our computer/laptop allows us to manage certain of its functionalities. Over slides and overheads, computer-based presentations have significant advantages. Audio, video, and even interactive programmes can be used to improve presentations. Unfortunately, employing these techniques is more complicated than using slides or overheads. The speaker must operate various devices with unfamiliar controls (e.g., keyboard, mouse, VCR remote control). In the dark, these devices are difficult to see, and manipulating them causes the presentation to be disrupted. Hand gestures are the most natural and effortless manner of communicating. The camera's output will be displayed on the monitor. The concept is to use a simple camera instead of a classic or standard mouse to control mouse cursor functions. The Virtual Mouse provides an infrastructure between the user and the system using only a camera. It allows users to interface with machines without the use of mechanical or physical devices, and even control mouse functionalities. This study presents a method for controlling the cursor's position without the need of any electronic equipment. While actions such as clicking and dragging things will be carried out using various hand gestures. As an input device, the suggested system will just require a webcam. The suggested system will require the use of OpenCV and Python as well as other tools. The camera's output will be presented on the system's screen so that the user can further calibrate it [11].

It is crucial to use hand gestures when communicating with one another: The task of utilizing hand gestures in technology is impacted by one of the most prevalent ways that people interact with their surroundings. Hand gesture analysis includes recognizing and position estimation of hands. Because the hand is smaller than other parts of the body, it is much harder to locate the pointing hand than other parts. The hand has greater complexity and more challenges due to differences between the cultural or individual factors of users and gestures invented from ad-hoc. The complications and divergences of finding hand gestures will deeply affect the recognition rate and accuracy. This paper emphasizes on summary of skeleton-based dynamic hand gesture technique, skeleton points extraction methods, merits and demerits, various applications, available data sets, and skeleton-based recognition methods. This paper also scrutinizes the performance of hand gesture recognition systems using skeleton data [12].

3. Methodology

3.1 Open CV

Open CV (Open-Source Computer Vision Library) is significantly aimed at real-time computer vision and its open library of python. It is also available in Java and C for making Open CV exercises. It covers the beginning of data structures which are similar as scalar, point etc. It is an open-source machine literacy software library. Open CV package is used in image and video processing and also analysis similar to face spotting and object spotting [13]. The Open CV package is imported into python using the decalogue 'import cv2'.

3.2 Architecture of Open CV

In computers, everything like videos, documents, images etc. are all converted and stored in the form of numbers. The pixel value will convert images into numbers. The pixel is the smallest unit of a digital image. By numbers, at a particular location, the intensity of the picture is represented. Open CV works in BGR format (Blue, Green, Red).

3.3 Mediapipe

MediaPipe is an open-source frame of Google and configuration which is used for applications in a

https://doi.org/10.36282/IJASRM/9.6.2024.1944

machine learning channel. Configuration of MediaPipe is grounded on three essential parts; they are performance evaluation, configuration for reacquiring detector data, that are operable, and a library of elements which are called calculators. MediaPipe uses a single-shot sensor model [14]. In real-time, for detecting and recognizing a hand or palm single- shot sensor model is used.

3.4 Architecture of Mediapipe

One crucial escalation Mediapipe provides is that the palm sensorial runs as required substantial calculation. We attain this by reasoning the position of the hand in the present video frames from the calculated hand landmarks in the former frame, barring the necessity to apply the palm sensor on every frame. For robustness, the hand tracker model also outputs another scalar capturing the confidence that a hand is present and considerately aligned in the input crop. For using multiple deep learning models Mediapipe Library is a real-time solution. Mediapipe implementation is very vast. There are two stages in detecting hand landmarks as shown in Fig. 1. First stage is the Palm Detector and the next stage is the Landmark Detector [15]. In the first frame of the module, the Bounding box around the Hand is detected by the Palm Detector. Then it is passed to the landmark detector module which detects key landmark points on the hand and allocates it. After the palm discovery over the whole image, the subsequent hand landmark model operates precise key point localization of 21 3-dimensional hand-knuckle coordinates inside the detected hand regions through regression, which is a direct coordinate prediction. The model learns a steady internal hand pose representation and is robust even to partly visible hands. In the subsequent frames, the same thing is repeated and the only difference is instead of calling the palm detection again and again we use a tracker. This saves lot of time. That is one of the reasons why MediaPipe is very fast [16].

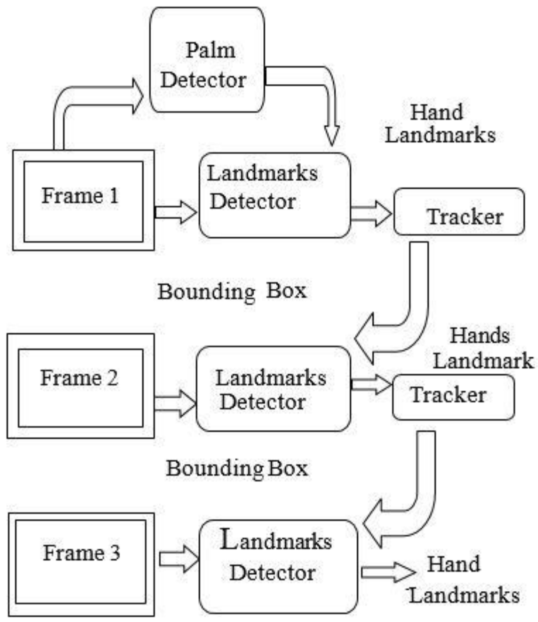

Fig. 1 Flow of the Mediapipe hand landmark detection.

3.5 Classifiers

A classifier is a type of machine learning algorithm used to specify class labels to data input in data science. Classifier algorithms employ sophisticated mathematical and statistical methods to generate predictions about the likelihood of a data input being classified in a given way [17]. The AI virtual mouse uses a Transformation algorithm. It transforms the coordinates of fingertips from the webcam screen to the computer window full screen for controlling the mouse.

3.6 Hand Tracking Module

The Hand Tracking Module consists of two main modules in the backend: Palm Detection and Hand Landmarks. The Palm Detection provides a cropped palm Image. Hand Landmarks find 21 different Landmarks as shown in Fig. 2. In this module Python Packages, Open CV and mediapipe are used. By creating this module, we can get the position values of the Hand Landmarks easily.

https://doi.org/10.36282/IJASRM/9.6.2024.1944

International Journal of Advanced Scientific Research and Management, Volume 9 Issue 5, June 2024

IJASRM

www.ijasrm.com

ISSN 2455-6378

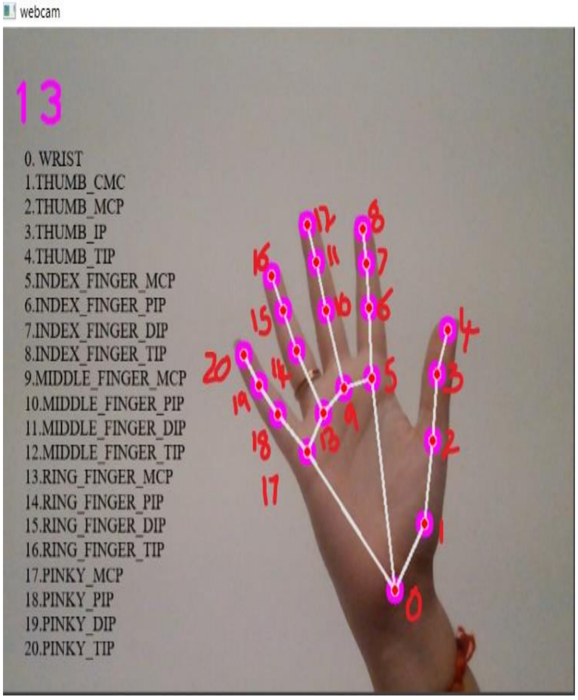

Fig. 2 Hand landmarks

3.6 System Implementation

The flow chart in Fig. 3 explains the various conditions and functions used in the AI virtual Mouse. The system is initialized and the video object is created using video capture. The webcam captures the RGB images of the Hand. The Hand and Hand Tips are detected using the mediapipe, Open CV packages. Hand Landmarks in Fig. 2 are highlighted and lines are drawn between the points in hand. Co-ordinates (Cx, Cy) of the points in Hand Landmarks are found. Rectangle Box is drawn which is the region of the PC window where we are going to use the mouse. In the next step, we are detecting which fingers are up using the Tip ID and the distance between two fingertips is also measured. Respective functions such as movement of the mouse cursor around the window and clicking operations are performed.

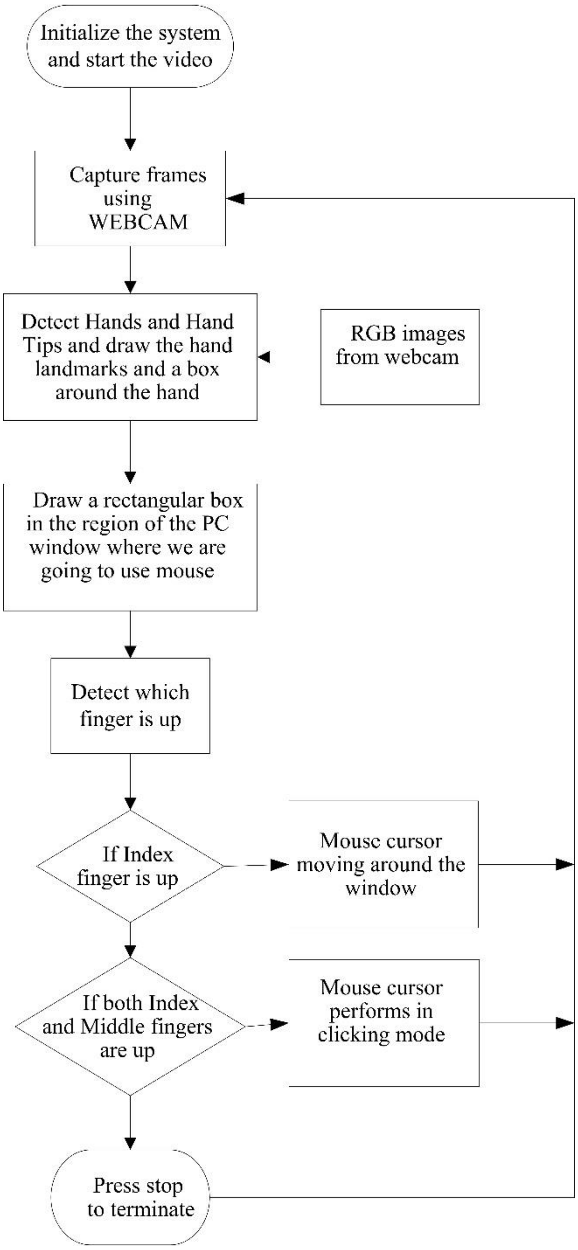

Fig. 3 Working of the AI Virtual Mouse

4. Results and Discussion

A process of presenting the on-screen cursor movement without any physical connection is discussed in this work. Spotting of the hand gestures and hand tips is done using a built-in camera or a webcam that processes these frames to perform the particular mouse operations. Different operations of mouse control like mouse movement and clicking operation are performed.

https://doi.org/10.36282/IJASRM/9.6.2024.1944

International Journal of Advanced Scientific Research and Management, Volume 9 Issue 5, June 2024

www.ijasrm.com

ISSN 2455-6378

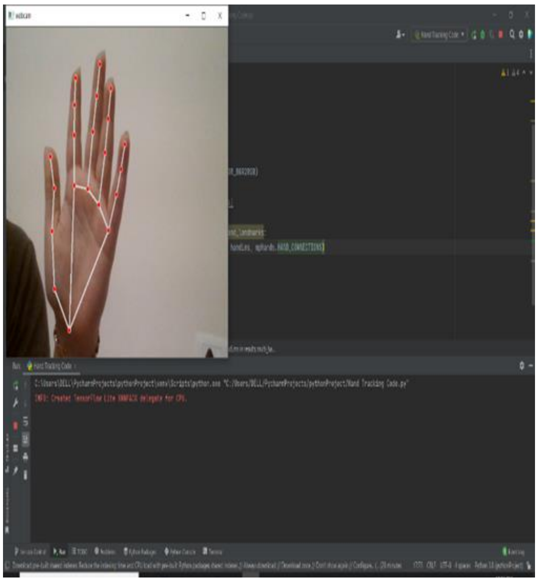

Fig. 4 Screen capture of the Hand Connections drawn using MediaPipe.

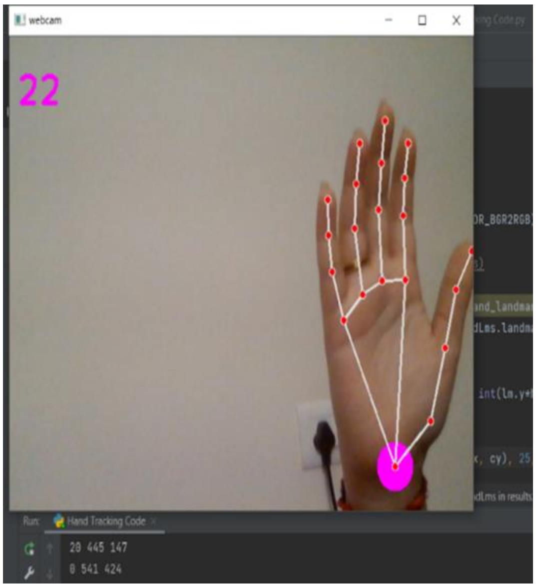

Fig. 5 Screen capture of the wrist landmark highlighted

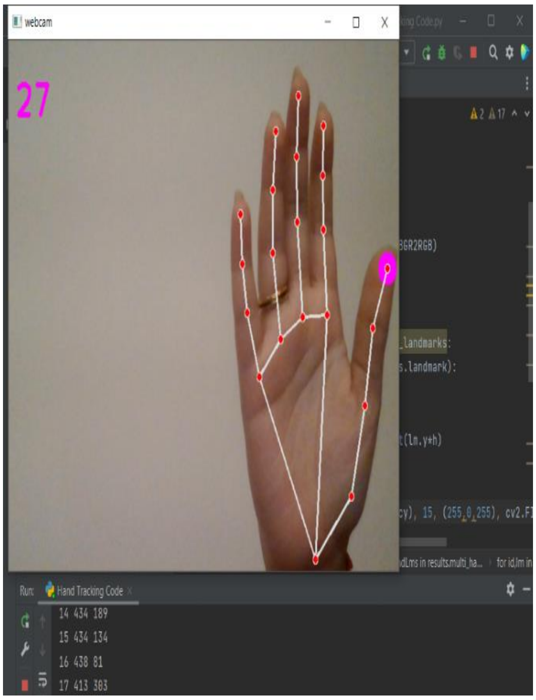

Fig. 6 Screen capture of the tip of thumb landmark highlighted

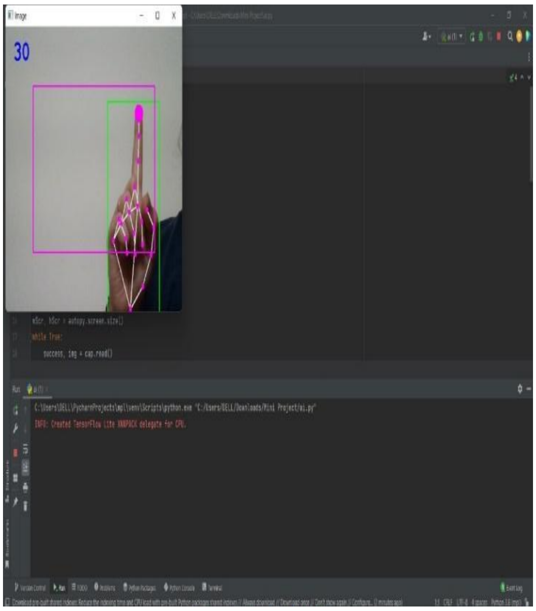

Fig. 7 Movement of the cursor

5. Conclusions and Future Scope

The basic motive is to control the cursor movement of the mouse with hand gesture instead of using a hardware mouse. The proposed work can be pursued by utilizing a built-in camera or webcam in order to perform certain mouse action that detects and interpret hand actions and hand tips. Rather than the standard physical mouse. The proposed work can be utilized virtually by using hand gesture. One main application

https://doi.org/10.36282/IJASRM/9.6.2024.1944

IJASRM
International Journal of Advanced Scientific Research and Management, Volume 9 Issue 5, June 2024
www.ijasrm.com
ISSN 2455-6378

in the real world is to decrease the spread of COVID-19.

In this AI Virtual Mouse, we were able to design a robust gesture recognition system which works well for the movement of the cursor and the clicking operation but there is still room for many improvements. This AI virtual mouse has some flaws like not being able to perform additional mouse functions such as right click, left click, increasing and decreasing the volume, dragging and dropping, and selecting text etc. In the future, these barriers can be overcome and can be cosigned. So future work will include the execution of additional gestures which will enable the user to perform more functions with ease. Aside from the above-mentioned, additionally, keyboard capability can be assimilated to imitate keyboard functions along with the mouse operations which shows to be the scope for the future.

Reference

[1] Sandeep Thakur, Rajesh Mehra, and Buddhi Prakash. "Vision based computer mouse control using hand gestures." 2015 International Conference on Soft Computing Techniques and Implementations (ICSCTI). IEEE, (2015).

[2] Haria A, Subramanian A, Asokkumar N, Poddar S, Nayak J S, "Hand gesture recognition for human computer interaction." Procedia computer science Vol. 115: Page No-367-374, (2017).

[3] Shajideen, Sherin Mohammed Sali, and V. H. Preetha. "Hand Gestures-Virtual Mouse for Human Computer Interaction." 2018 International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE, (2018).

[4] Varun, Kollipara Sai, I. Puneeth, and T. Prem Jacob. "Virtual Mouse Implementation using OpenCV." 3rd International Conference on Trends in Electronics and Informatics (ICOEI). IEEE, (2019).

[5] Shriram S, Nagaraj B, Jaya J, Shankar S, Ajay P. "Deep Learning-Based Real-Time AI Virtual Mouse System Using Computer Vision to Avoid COVID-19 Spread." Journal of Healthcare Engineering (2021).

[6] Orkun Ozturk, Alper Aksac, Tansel Ozyer and Reda Alhajj, "Boosting real-time recognition of hand posture and gesture for virtual mouse operations with segmentation." Applied Intelligence Vol. 43, (No. 4): Page No-786-801, (2015).

[7] Zhang F, Bazarevsky V, Vakunov A, Tkachenka A, Sung G, Chang C L, Grundmann M. "Mediapipe hands: On-device real-time hand tracking." arXiv preprintarXiv:2006.10214 (2020).

[8] Julian, Anitha, Danush Suresh, and R. Ramyadevi. "Gesture-Driven Virtual Mouse: Empowering Accessibility through Hand Movements." IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT). Vol. 5. IEEE, (2024).

[9] Dharavath, Krishna, "Gesture and Voice Controlled Virtual Mouse for Elderly People." 2nd International Conference on Networking and Communications (ICNWC). IEEE, (2024).

[10] Gunda, Chidvika, Manohar Maddelabanda, and Hariharan Shanmugasundaram. "Free hand text displaying through hand gestures using mediapipe." 3rd International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT). IEEE, (2022).

[11] Matlani and Roshnee, "Virtual mouse using hand gestures." International conference on technological advancements and innovations (ICTAI). IEEE, (2021).

[12] Tripathi, Reena and Bindu Verma. "Skeleton Data is all about: Dynamic Hand Gesture Recognition." 7th International Conference on Image Information Processing (ICIIP). IEEE, (2023).

[13] Chowdhury, Sugnik Roy, Sumit Pathak, and M D Anto Praveena. "Gesture Recognition Based Virtual Mouse and Keyboard." 4th International Conference on Trends in Electronics and Informatics (ICOEI) (48184). IEEE, (2020).

[14] Katona J, A Review of Human–Computer Interaction and Virtual Reality Research Fields in Cognitive Info Communications, “Applied Sciences”. Vol. 11, (No. 6): Page No-2646, (2021).

[15] Varun, Kollipara Sai, I. Puneeth, and T. Prem Jacob. "Virtual Mouse Implementation using OpenCV." 3rd International Conference on Trends in Electronics and Informatics (ICOEI). IEEE, (2019).

[16] Reddy V V, Dhyanchand T, Krishna G V, and Maheshwaram S. "Virtual Mouse Control Using Colored Finger Tips and Hand Gesture Recognition." IEEE-HYDCON. IEEE, (2020).

[17] Tran D S, Ho N H, Yang H J, Kim S H, and Lee G S. "Real-time virtual mouse system using RGB-D images and fingertip detection." Multimedia Tools and Applications Vol. 80, (No. 7): Page No-10473-10490, (2021).

https://doi.org/10.36282/IJASRM/9.6.2024.1944