Painting with Hand Gestures using MediaPipe

R.Vasavi[1]
Assistance Professor
Department of CSE VNR VJIET Hyderabad, India

Nenavath Rahul[2], A Snigdha[3], K. Jeffery Moses[4], S.Vishal Simha[5]
Department of CSE VNR VJIET Hyderabad, India

Abstract:- The main objective of this project is that the hand gesture recognition can also be utilised in applications including industrial automation control, sign language interpretation, and rehabilitation equipment for individuals with physical disabilities of the upper extremities. And it is also find applications in varied domains like virtual environments, medical systems, smart surveillance etc. Hand gesture recognition is most significant for human-computer interaction. Gesture Recognition is a technique that uses mathematical algorithms to recognise human gestures. Gesture recognition identifies the hand, monitors hand motions, and also reports on the hand's orientation and finger flow. The user's fingers are tipped with colour markers. The color markers are placed at the tip of the user fingers. The drawing application allows the user to draw by tracking the fingertip movements of the user's index finger. The existing system has hand gesture recognition serves as a key for overcoming many difficulties and providing convenience for human life. This number recognition through hand gestures project recognizes the numbers from 0-9 according to the gestures of hand.

Keywords:- *Hand Gesture Recognition, Computer Vision, Paint, Media Pipe, OpenCV.*

I. INTRODUCTION

The mouse, keyboard, remote control, touch screen, and other similar devices are currently the main ways in which people and machines interact directly. People communicate primarily through non-contact methods that are natural and intuitive, such as sound and physical motion. Many researchers have attempted to assist computers in identifying other people's intentions and information by using non-contact methods such as voice, facial expressions, physical motions, and gestures. Gestures are the most important aspect of mortal language, and they also play an important role in mortal communication. They are thought to be the most basic way for humans and computers to communicate. Gesture recognition encompasses sign language recognition, robotics, and other applications.

Gesture recognition for HCI applications typically employs two methods. The first relies on wearable or direct physical approaches, whereas the second relies on computer vision and does not require the use of any sensors. In the wearable or direct contact technique, the data-glove, which is made up of sensors to capture hand motion and location, is used. The camera is used in the vision-based technique to provide contactless communication between humans and machines. Computer vision cameras are simple to use and reasonably priced. Due to variations in hand sizes, positions, and orientations as well as in lighting and other factors, this approach does have some limitations.

In this paper, we present a painting technique that uses hand motions to sketch or draw in real time on the canvas. Utilizing cameras to record hand motion, hand gesture-based paint applications can be deployed. Using vision-based real-time dynamic hand movements, an intangible interface is developed and put into use to carry out tasks including tool selection, canvas writing, and canvas clearing. The system's web camera captures images of the hands, which are then analyzed in real time with a single-shot detector model and media pipe to enable the machine to communicate with the user in a split second.

II. RELATED WORK

To understand all the prior research on the project that we have taken on, we read a wide variety of research papers.

- Since the computer vision and machine learning software library OpenCV is free and open-source, we have learned about its benefits in terms of how well it captures images and videos.To facilitate the use of machine perception in commercial goods and to provide a common foundation for computer vision applications, OpenCV was developed. A package called OpenCV is used to carry out tasks like face detection image processing. Analyze the video by tracking objects, estimating the motion, and removing the backdrop. Computer vision is a field that describes how to create, alter, and comprehend a 3D scene from its 2D images based on the characteristics of the scene's existing structure. It focuses on using computer hardware and software to simulate and replicate human vision.
- Image processing it covers a variety of image processing techniques, including histograms, color space conversion, geometric picture modifications, and image filtering. This module is a component of the OpenCV Java library package org.opencv.imgproc.
- Video analysis like object tracking, background subtraction, and motion estimation are covered in this particular module. This module is included as a package name org.opencv.video in the OpenCV Java library. [4]Core Functionality's the data structures used to create OpenCV applications, Point, including Scalar, Range, etc., are covered in this module. In addition to these, it also includes the multidimensional array Mat, which is used to store the images. The OpenCV Java library includes this module as part of a package with the name org.opencv.core.

- Effective resource management is key to achieving low latency performance. MediaPipe also handles the synchronisation of time series data, such as audio and video frames, and serves other essential functions. MediaPipe abstracts each perception model into a module and connects it to a maintenance graph. The TensorFlow and TF Lite inference engines are supported by MediaPipe in addition to the features mentioned above. Any TensorFlow and TF Lite model may be used on MediaPipe. At the same time, MediaPipe offers GPU acceleration of the device itself on mobile and embedded platforms. tags, but they are used to interpret the page's content.
- Sayem Mohammad Siam, Md. Hasanul Kabir, and Jahidul Adnan Sakelhas proposed a new method of HCI that uses marker detection, tracking technique. Instead of having a mouse or touchpad, two colored markers are worn on the tips of the fingers to generate the eight hand movements to provide instructions to the laptop or desktop computer with a

consumer-grade camera. They also utilized the "Template matching" technique to detect markers and the Kalman Filter to track them. To recognize real-time dynamic hand motions, the developed system employs a data glove-based technique. The data glove contains ten soft sensors that monitor the joint angles of five fingers and gather gesture data. Techniques such as gesture spotting, gesture sequence simplification, and gesture recognition are used to recognize real-time motions.

- Shomi Khan, M. Elieas Ali, Sree Sourav Das created a system which converts ASL (American Sign Language) into the text from real-time video using a skin color identification algorithm. Because hand shape and the skin of the person vary from person to person, distinguishing the hand may be difficult. To solve this, the system employs two neural networks. The first is the SCD (Scalable color descriptor) neural network. The image pixels are put into the SCD neural network, which decides if they are skin pixels or not. The second is a neural network called HGR (Hand gesture recognition), to which the retrieved features will be fed. Two independent methods, Finding the fingertip and Pixel segmentation, will be used to extract the characteristics.
- By just sweeping their finger over a vibrant LED light source, Pavitra Ramasamy and Prabhu G have proposed a new technology that would allow users to create the alphabet or type anything they desire. To extract the movement of the finger sketching the alphabet, only the color of the LED is tracked. The tracked object's color is changed to white, and the background is made of black. The user wanted to draw an image of the alphabet in black and white, so they stitched together several black and white frames to make it.
- To enable mouse functions like dragging the mouse cursor and clicking left, right, and centre with hand gestures, an intangible interface is created and put into use using vision-based real-time dynamic hand gestures. The system is put into action using MATLAB. A new HCI method created by S. Belgamwar and S. Agrawal includes two Arduino microcontrollers, an accelerometer, a camera, and ultrasonic distance sensors. The main idea behind this interface is to use the ultrasonic distance sensors to collect motions. To record the gestures, the distance between the sensor and the hand is determined.
- Hazem Wannous, Jean-Philippe Vandeborre, and Quentin De Smedtused a skeleton-based model for 3D hand motion identification. They extracted an efficient descriptor from the hand skeleton connected joints of the Intel Real-Sense depth camera using the geometric shape of the hand. The skeleton-based method outperforms the depth-based method. RevatiKhadse and Saina Rasal have created the virtual paint programme Prajakta Vidhate, which tracks hand motions and writes on the screen using ball-tracking technology. As a contour, they've utilized a glove with a ping pong ball fastened to it. A configurable airbrush model that makes use of the Leap Motion Controller, which can track hands, to produce an immersive freehand painting experience was demonstrated byWei Chen, Fei Chen, Yuefeng Ze, and RuiminLyu.

III. PROPOSED SYSTEM

The proposed system is an application, in which the user can paint using his fingertip. Unlike paint, no need to drag or draw it recognizes the hand movement in the direction of a needed diagram. By using OpenCV we paint on the canvas and MediaPipe for the hand-tracking.

This model recognizes the hand and the fingertip of the index finger with the help of the MediaPipe framework which is provided by Google. The Media Pipe framework consists of many inbuilt Machine Learning solutions and OpenCV for capturing the movements of the hand and tip of the index finger according to the movement of the tip the user can paint on the canvas.

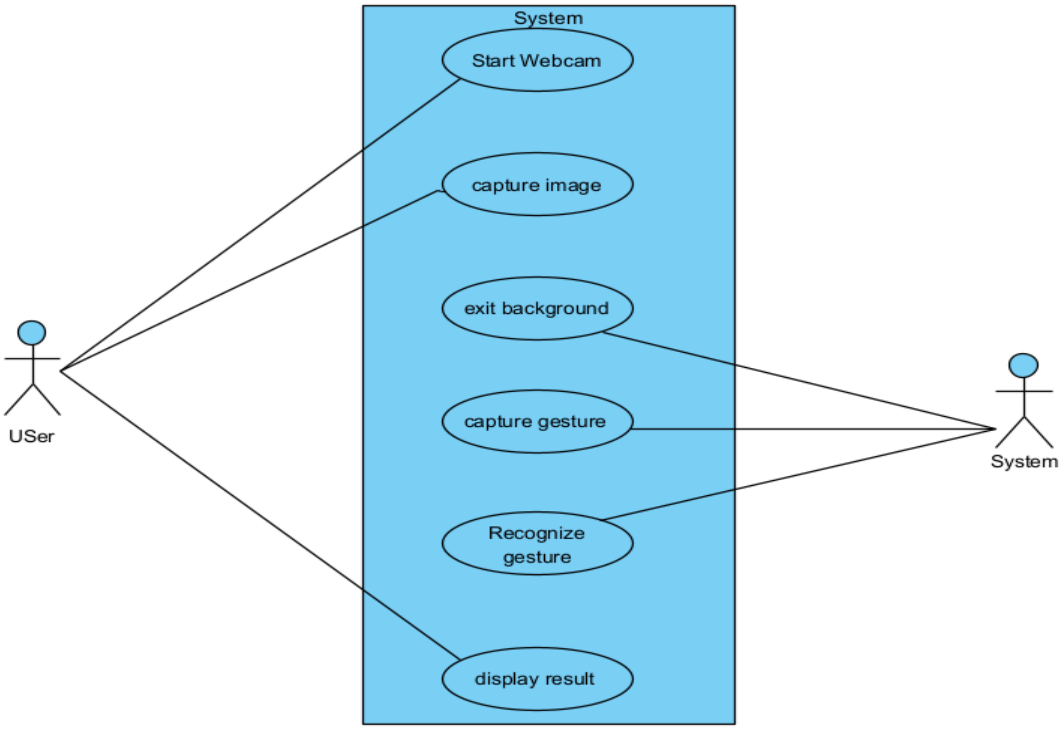


Fig. 1: Basic Flow of Application

IV. ALGORITHM AND LIBRARIES

The MediaPipe framework handles hand gesture identification and tracking, and the OpenCV library handles computer vision. The application uses machine learning concepts to track and identify hand motions and hand tips.

The MediaPipe is the open-source framework it consist on many pre-built machine learning solutions with the help of MediaPipe framework as it has the solutions in it. The project is using all including the solutions which ever are needed to solve the problem and with help of it for the Palm detection it uses Palm Detection Model and the Hand Landmark Model. With the Help of the Hand Landmark Model it could detect the index finger tip.

A. MediaPipe

A Google open-source framework called MediaPipe was first made available in 2019. Machine learning and computer vision features are included into MediaPipe. Using MediaPipe, a machine learning inference pipeline is put into action. Running real data points using ML inference is the method. The MediaPipe framework is used to address AI problems, the majority of which involve streaming audio and video. It's cross-platform and multimodal. The framework is used to produce cross-platform applications as a consequence. A few of the applications provided by MediaPipe includehair segmentation, face detection, multi-hand tracking, object detection, and monitoring. A framework with a high degree of fidelity is MediaPipe. The MediaPipe framework offers low latency performance. Time-series data synchronization is its responsibility. The MediaPipe framework is used to design and analyze systems, as well as to develop the systems for application purposes.

There are several good reasons why MediaPipe is required, including managing resources (CPU and GPU) efficiently to achieve low latency performance, handling the synchronisation of time series data such as audio and video frames, and more. Each perceptual model is turned into a module by MediaPipe, which connects them with a maintenance graph. In additional to the functionalities listed above, MediaPipe supports the inference engines TensorFlow and TF Lite. On MediaPipe, any TensorFlow and TF Lite model may be applied. MediaPipe simultaneously offers GPU device acceleration on both mobile and embedded devices. tags, but interpret the information they contain.

Google has created the open-source Mediapipe framework, which is still in alpha, to "develop word-class solutions for machine learning. Although it has been publicly available for a year, it has probably been in development for much longer. The fact that Mediapipe's code is written in C++ but can easily be deployed to any platform—from web assembly to Android to MacOS—is a strong "selling" element (besides the fact that it's free).

When Mediapipe was first launched, it only offered a few demos. Now, however, its GitHub website offers around a dozen distinct demos, ranging from pose tracking that may overtake OpenPose to persistent object tracking and AR hair colouring.

ISSN No:-2456-2165

B. *MediaPipe is Fast*

Our team had two possibilities while looking for hand tracking software: OpenPose and MediaPipe. The extreme disparity in pace between the two was what first caught my attention. On a slow Mac, MediaPipe could only manage 7 frames per second, while OpenPose would be fortunate to manage 7 seconds. GPU acceleration and multi-threading are used by MediaPipe to attain its speed. As long as you use excellent graph-making skills, MediaPipe takes the reins and handles these typically challenging development techniques for you. Modern smartphones can run away with frames thanks to multi-threading and GPU acceleration, frequently at FPS levels that are too high for human vision. However, don't anticipate your Motorola from 2015 to keep up.The fact is that machine learning is computationally ten or twenty times more complicated than Flappy Bird.

C. *Mediapipe is Modular and Reusable*

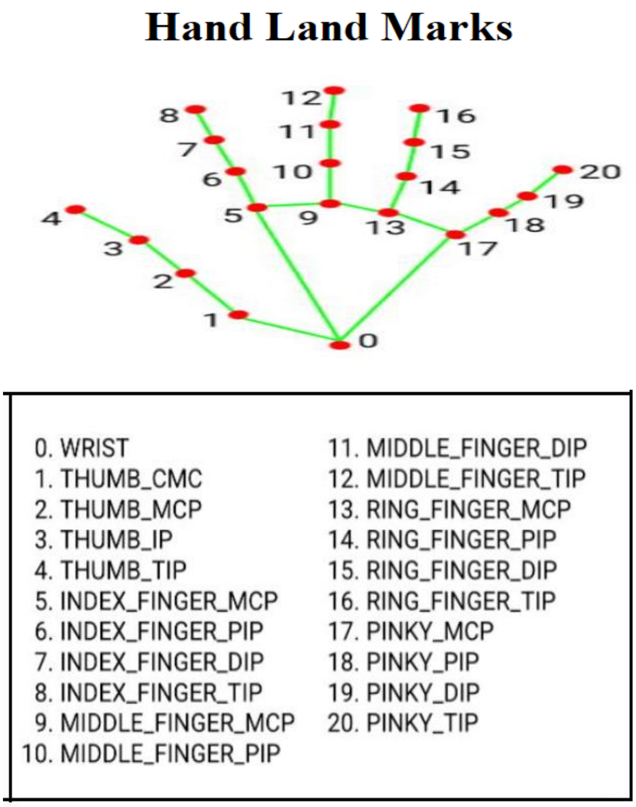
The use of graphs, subgraphs, and calculators by Mediapipe makes it simple to transfer the work of one project to another. When you add side packets to that, you can actually adjust each calculator's settings to meet various projects.

You can freely use the numerous "sample calculators" that Mediapipe already has, such as multi-platform renderers, multi-platform TensorFlow Lite, and pre-made neural networks.A small development team may find it challenging to support multiple operating systems, including Windows, Mac, and Ubuntu.

However, you can deploy your application on mobile devices in addition to PCs in a matter of hours (or minutes, if you're experienced). Web assembly support is another feature that MediaPipe touts, but our team was unable to test it. Take into account the fact that MediaPipe lacks Windows support while developing desktop apps..

Currently, MediaPipe is the best option for deploying neural networks to mobile devices, and it also works reasonably well for deploying neural networks to desktop computers. I wouldn't have been on team MediaPipe for desktop without Windows support, however keep in mind that Web Assembly is an alternative in that scenario. However, users should be aware that since MediaPipe is still in alpha, there is a lack of documentation and changes may occur quickly, disrupting your workflow. Some people might not consider this a disadvantage because less of their rivals may be prepared to utilise MediaPipe in its current form, which would give them a significant advantage when MediaPipe version 1.0 is released.

The pipeline arrangement, each and every steps of the system are executed. The pipeline, which built to scale and be able to run on a variety of systems. Between desktop and mobile platforms. Performance evaluation, obtaining sensor data, and gathering Each component is included in the MediaPipe framework. The components of the system are calculators. A Mediapipe Real-time framework utilizes a single-shot detector model for identifying and detecting a hand or palm. First, it is trained regarding the hand detection module's palm detection model so training palms is simpler. It refers to a hand landmark in the hand area with 21 joints or knuckle coordinates as in fig-2.



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

Fig. 2: Hand Land Marks

D. OpenCV

The OpenCV computer vision library is a must-have for anyone who tends to work with computers. It includes image-processing methods for object detection. Real-time computer vision applications can be created using the Python library OpenCV. The processing and analysis of images and videos is handled by the OpenCV library.

In terms of the characteristics of the structure existing in the scene, computer vision can be defined as a discipline that describes how to reconstruct, disrupt, and comprehend a 3D scene from its 2D images. It focuses on simulating and reproducing human vision with computer hardware and software.

Significant crossover exists between the following fields and computer vision.−

- Image Processing − Image modification is the main subject.
- Pattern Recognition − It describes several methods for categorising patterns.
- Photogrammetry − This field focuses on extracting precise measurements from photographs.

a) Image Processing
This subject covers a variety of image processing techniques, including histograms, colour space conversion, geometric picture modifications, and image filtering. This module is included as a package with the name org.opencv.imgproc in the OpenCV Java library.

b) Video
Concepts for video analysis like object tracking, background subtraction, and motion estimation are covered in this module. This module is included as a package with the name org.opencv.video in the OpenCV Java library.

c) Core Functionality
The fundamental data structures necessary to create OpenCV applications, including Scalar, Point, Range, etc., are covered in this topic. It also contains the multidimensional array Mat, which is used to hold the photos, in addition to these. This module is included as a package with the name org.opencv.core in the OpenCV Java library.

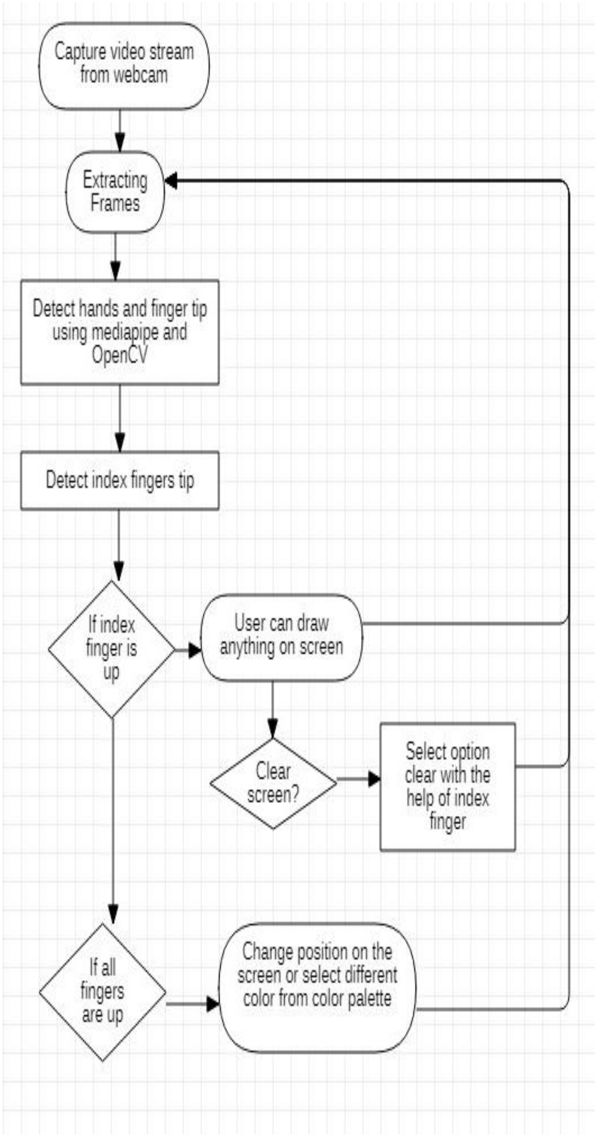

Capture video stream from webcam
Extracting Frames
Detect hands and finger tip using mediapipe and OpenCV
Detect index fingers tip
If index finger is up
User can draw anything on screen
Clear screen?
Select option clear with the help of index finger
If all fingers are up
Change position on the screen or select different color from color palette

Fig. 3: Flowchart of painting with hand gestures

The painting with hand gestures application is based on the number of webcam frames recorded. The webcam captures the frames and sends them to the system. It captured each frame with the web camera until the application was decided to close. The frame number of interest we need to set the position we need using set and extract the frame using read. The read function returns the frames itself and the other value ret. If the frame is found then it will be True or else it will be False. The system will determine which finger is up by comparing the tips of the corresponding fingers explored using MediaPipe to the relevant coordinates of the fingers and then performing the actions accordingly. If the index finger is raised, the user can draw anything on the canvas. If all of the fingers are up, the user can move around on the screen. The user can also use the index finger to select any colour scheme or clear the screen.

V. RESULTS

This application is used to paint with hand gestures. The System can detect the live video of a person and recognizes the index fingertip, based on the hand gesture and movement the person can draw and view what they wanted to do like painting, writing some text. The System can display the paint which the person has written using free hand gestures.

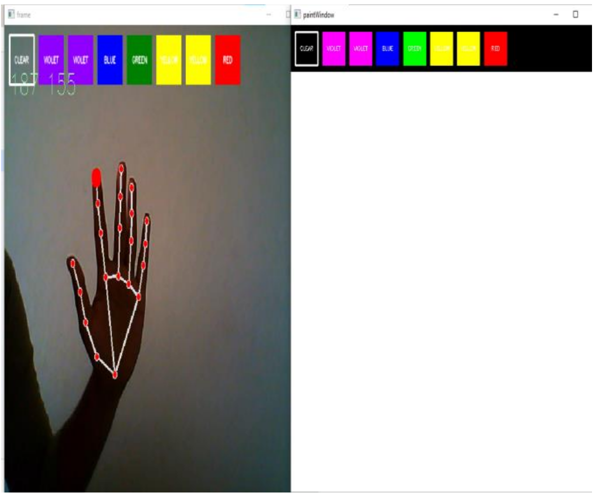


Fig. 4: This image shows the empty canvas.

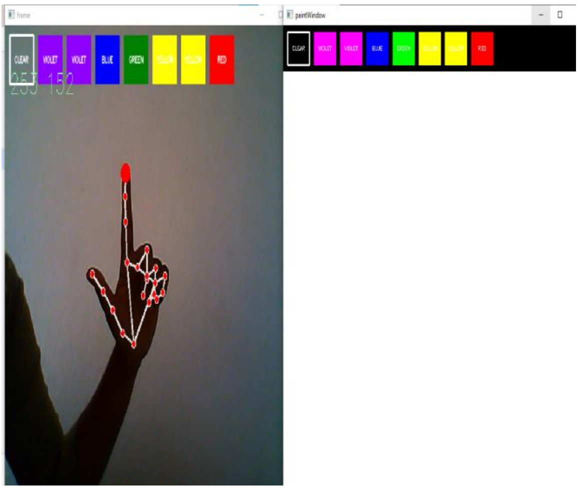


Fig. 5: This image shows pen up page(no operations will be performed)

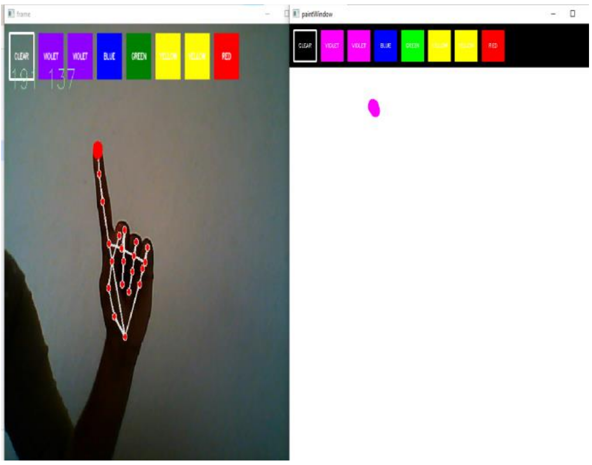


Fig. 6: The image shows pen down page

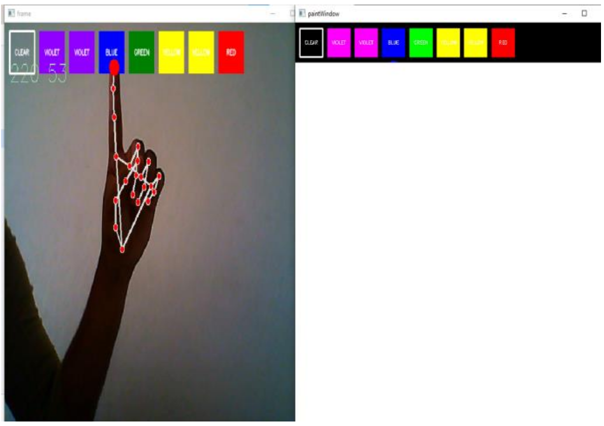


Fig. 7: This image shows selection of desired color.

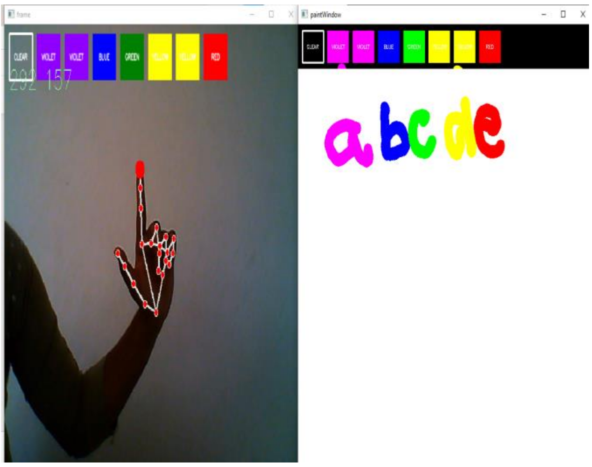


Fig. 8: This image shows painting or witing

VI. CONCLUSION

The painting with hand gestures applications main goal is to deliver the user a AI-based application that allows the used to draw anything on canvas with the free hand using the index fingertip. The application also provieds the option for the user to choose the color of his choose to draw on the canvas.

REFERENCES

[1.] Saurabh uday Saoji, Bharati Vidyapeeth, Akash Kumar Choudhary and Bharat Phogat – “Basic Paint Window Application via Webcam Using OpenCV and Numpy in Python”, August 2021

[2.] Ayesha Gurnani, Viraj Mavani, and Vandit Gajjar – “Hand gesture real time paint tool-box: Machine learning approach”, 2018, DOI: 10.1109/ICPCSI.2017.8391833 – IEEE Publisher

[3.] Kandarpa Kumar Sarma, Dharani Mazumdar, and Anjan Kumar Talukdar - “Gloved and Free Hand Tracking based Hand Gesture Recognition”, September 2013, DOI: 10.1109/ICETACS.2013.6691422 - IEEE Publisher

[4.] Janaka Wijayanayaka and Matheesha Fernando-“Low cost approach for Real Time Sign Language Recognition”, February 2014, DOI: 10.1109/ICIInfS.2013.6732059 - IEEE Publisher

[5.] Chung-Chieh Kao, Yo-Jen Tu, and Huei-Yung Lin - "Human Computer Interaction Using Face and Gesture Recognition", January 2014, DOI: 10.1109/APSIPA.2013.6694276 - IEEE Publisher

[6.] Mithila Hadap, Rucha Khatal, Sanjana Khupase, Surbhi Asati, and Archana S. Ghotkar - "Hand Gesture

[7.] Recognition for Indian Sign Language", March 2012, DOI: 10.1109/ICCCI.2012.6158807 - IEEE Publisher

[8.] Bodo Rosenhahn, Laura Leal-Taixé, and Alina Kuznetsova -"Real Time Sign Language Recognition using consumer depth camera", March 2014, DOI: 10.1109/ICCVW.2013.18 - IEEE Publisher

[9.] Bo Yuan and Yanmin Zhu, Zhibo Yang- "Vision Based Hand Gesture Recognition", May 2013, DOI: 10.1109/ICSS.2013.40 - IEEE Publisher

[10.] Suree Pumrin andThittaporn Ganokratanaa - "The Vision-Based Hand Gesture Recognition Using Blob Analysis", April 2017, DOI: 10.1109/ICDAMT.2017.7904987 - IEEE Publisher

[11.] Gaurav Rokade, Pradeep Kurund, Ajay Ahire, Prashant Bhagat and Vishnu Kamble - "Paint using Hand Gesture Recognition for Human Computer Interaction", February 2020 – IRJET

[12.] Pankaj K. Bharne and Shweta K. Yewale - "Hand Gesture Recognition Using Different Algorithms Based on Artificial Neural Network", August 2012, DOI: 10.1109/ETNCC.2011.6255906 - IEEE Publisher

[13.] Jeevanshi Sharma1 , Rajat Maheshwari2 , Salman Khan3 , Abid Ali Khan4 - "Evaluating Performance of Different Machine Learning Algorithms for the Acute EMG Hand Gesture Datasets.", Journal of Electronics 4, no. 3 (2022)

[14.] M. J. Cheok, Z. Omar, and M. H. Jaward, "A review of hand gesture and sign language recognition techniques," International Journal of Machine Learning and Cybernetics, vol. 10, no. 1, pp. 131–153, 2019

[15.] Cheok, Ming Jin, Zaid Omar, and Mohamed Hisham Jaward. "A review of hand gesture and sign language recognition techniques." International Journal of Machine Learning and Cybernetics 10, no. 1 (2019): 131-153.

[16.] V Shiva Narayana Reddy 1 , K. Kavya2 , D Sudheer Reddy3 – "Virtual Paint And Volume Control Using Hand Gesture" - Journal of Positive School Psychology, 2022, Vol. 6, No. 8, 153-159