IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

AIR CANVAS USING OPENCV AND MEDIAPIPE

[1]NEERAJ RAJESH JAVALAGI, [2]SAHIL PATIL, [3]YASH GUNJAL [4]SHANTANU VAIDYA

[5]TANMAYEE KUTE

[1]Student, [2]Student, [3]Student, [4]Student, [5]Professor
[1]Information Technology Department,
[1,2,3,4,5]PROGRESSIVE EDUCATION SOCIETY'S MODERN COLLEGE OF ENGINEERING, PUNE, INDIA

Abstract: The concept of writing in the air, a subject of growing fascination and significance in the realms of image processing and pattern recognition, offers exciting potential for educational applications. Recent research has been dedicated to developing innovative techniques that expedite the processing of information and enhance the accuracy of recognition, making it particularly valuable for educational purposes. In the field of Computer Vision, object tracking plays a crucial role and holds substantial educational value. Object tracking algorithms find applications in areas such as facilitating interactive learning experiences. The project is strategically positioned to address an essential educational need by concentrating on the creation of a motion-to-text converter. This technology's application serves as an asset in educational fields, allowing users to transcribe input gestures into image data using Computer Vision.

Index Terms – Computer Vision, Open CV, Media Pipe

I. INTRODUCTION

In the realm of education, the art of teaching is undergoing a significant transformation, with traditional methods gradually making room for the innovative potential of digital tools like Air Canvas. This platform leverages the power of hand gesture recognition, computer vision, and Python programming to create an interactive and immersive learning experience for students.

By incorporating Air Canvas into the classroom, educators can revolutionize the way students grasp various concepts, enabling a seamless fusion of digital and traditional learning methods. This technology opens new avenues for creative expression, allowing students to explore complex ideas through visual and interactive demonstrations. Whether it's illustrating scientific processes, exploring historical events, or understanding abstract concepts, Air Canvas serves as a powerful tool for enhancing comprehension and engagement.

Through its intuitive interface, Air Canvas bridges the gap between traditional and digital educational approaches, fostering a natural and dynamic interaction between students and the learning material. By utilizing diverse writing techniques such as digital pens, touch-screen surfaces, and gesture recognition, this platform encourages a hands-on and immersive learning environment, nurturing students 's creativity and critical thinking skills.

As the demand for modern educational solutions continues to grow, Air Canvas represents a transformative step towards fostering a deeper understanding of various subjects. By infusing the art of writing with digital innovation, this platform empowers educators to create an enriching and dynamic educational experience that cultivates a lifelong love for learning and exploration among students.

II. RELATED WORK

2.1 In 2020, Alamat et al. introduced a trajectory-based writing system that leverages three-dimensional trajectories recorded by a depth camera tracking fingertip movement. The recognition model they devised integrates Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) architectures. This model exhibited exceptional performance, achieving an impressive 99.32% accuracy on a 6D motion gesture alphanumeric character dataset, marking it as the highest accuracy reported to date. The experimental setup involved utilizing an IntelSense SR300 camera in conjunction with a computer. Programming interfaces were established using both C# and Python languages to facilitate seamless interaction with the hardware and data processing. For the implementation of the recognition model, Keras, a high-level API for TensorFlow, served as the backbone. A noteworthy aspect of the study was the consideration of varying ending frames for each character, attributable to distinct individual writing patterns. Despite this variability, the LSTM-CNN model demonstrated robust performance, showcasing its efficacy in recognizing and accurately classifying alphanumeric characters based on 6D motion gestures.

2.2 In 2021, Nishta Dua et al. introduced a groundbreaking project called the "motion-text-converter." The core concept of this system revolves around employing a set of gestures, equal in number to the functionalities they control. The fundamental functionalities include Writing mode, Colour mode, and Backspace. The primary input for this system is a video clip capturing hand movement. The data collection process involved capturing two-second videos of hand movements in diverse environmental settings. These videos were subsequently segmented into 30 individual images each, resulting in a dataset comprising approximately 2000 images. To facilitate model training and evaluation, the dataset was partitioned into training and development sets, with an 85%-15% split. For the implementation of their system, the authors utilized pre-trained models, specifically SSD (Single Shot Multibox Detector) and Faster-RCNN (Region-based Convolutional Neural Network). Notably, the Faster-RCNN model outperformed SSD, yielding significantly better results in terms of accuracy. Remarkably, the developed model achieved an impressive accuracy rate of 99%. However, it's important to note that this high accuracy was achieved under the condition that the generated images originated from the same video source. Consequently, the model exhibited limitations when presented with discrete backgrounds, indicating a dependency on the consistency of the video source for optimal performance. Despite this constraint, the motion-text-converter project represents a significant advancement in gesture-based control systems, showcasing exceptional accuracy and functionality in controlled environments with consistent backgrounds.

2.3 In 2022, M.Bhargavi, introduced a project**,** The Air Canvas project presents a revolutionary solution to the challenges associated with conventional drawing systems, particularly the limitations of mouse-based interfaces. By leveraging accessible methods and libraries such as Mediapipe, this project offers an efficient alternative, allowing users to effortlessly translate their imagination into digital drawings by simply waving their hands.The incorporation of Mediapipe proves instrumental in streamlining the process of hand position tracking. This not only enhances the efficiency of the air canvas system but also simplifies the intricate task of finger position detection, mitigating the need for complex image processing procedures. The versatility of this system extends its applicability to various domains, including education and artistic endeavors. By eliminating the reliance on hardware components like mice and touch screens, the air canvas project introduces a user-friendly and intuitive drawing experience. Moreover, it serves as a foundational framework for diverse applications requiring hand tracking capabilities.The project's potential impact on creativity is noteworthy, providing a novel and interactive platform for teaching and drawing. This innovative approach facilitates a more natural and engaging interaction compared to traditional methods. Looking forward, the Air Canvas project could serve as a cornerstone for future developments in hand tracking applications. Possible extensions include applications in sign language detection, virtual mouse functionalities, and other interactive systems that benefit from precise hand tracking. The demonstrated versatility positions this project as a catalyst for advancements in human-computer interaction, with the potential to reshape the landscape of various interactive technologies.

III. METHODOLOGY

3.1 OpenCV Library

OpenCV (Open - Source Computer Vision Library) is an open-source computer vision and machine learning software library. It provides a wide range of tools and functions for image and video analysis, including various image processing techniques, feature detection, object recognition, and more. OpenCV is widely used in computer vision applications, and it has bindings for multiple programming languages, including C++, Python, and Java.

The use of OpenCV in our project:

1. Gesture Recognition: OpenCV is used to capture video frames from a camera and process them for gesture recognition. Techniques such as background subtraction, contour detection, and convex hull can be employed to identify the hand or a specific object in the video stream.

2. Hand Tracking: Once the hand is identified, OpenCV can be used to track its movement across consecutive frames. This allows you to understand the trajectory of the hand in the air.

3. Drawing on Canvas: OpenCV can be integrated with a drawing canvas. As the hand moves in the air, you can use its position to draw lines, shapes, or patterns on the canvas. The coordinates of the hand in each frame can be translated into drawing commands on the canvas.

4. Colour Detection: OpenCV is used for colour detection. By tracking an object, you can map its position to draw in different colors.

5. User Interface: OpenCV is integrated with a graphical user interface (GUI) library to create a user-friendly interface for the air canvas. This will include features like selecting colours, clearing the canvas, or saving the drawings.

6. Calibration: To improve accuracy, you can implement a calibration step where the system learns the characteristics of the user's hand movements. This can involve capturing some initial data to set a baseline for hand tracking.

7. Noise Reduction: OpenCV provides tools for noise reduction in images. This can be useful in cleaning up the input from the camera, reducing jitter in the hand tracking, and improving the overall drawing experience.

3.2 MEDIAPIPE Library

MediaPipe is an open-source framework developed by Google that provides a set of machine learning (ML) solutions for various tasks related to computer vision and multimedia. One of its modules, known as MediaPipe Hands, can be utilized for implementing the AIR CANVAS project. It enables building machine learning pipelines for processing time-series data like video, audio, etc. This is a cross-platform Framework which works on Desktop/Server, Android, iOS, and embedded devices like Raspberry Pi and Jetson Nano.

3.3 Methodology of the project

Figure 1: State Chart Diagram represents the different states of an object during its lifetime. In the context of the Air Canvas project, which aims to facilitate teaching various concepts to students through digital art, a state chart diagram can help illustrate the different states and transitions within the application.

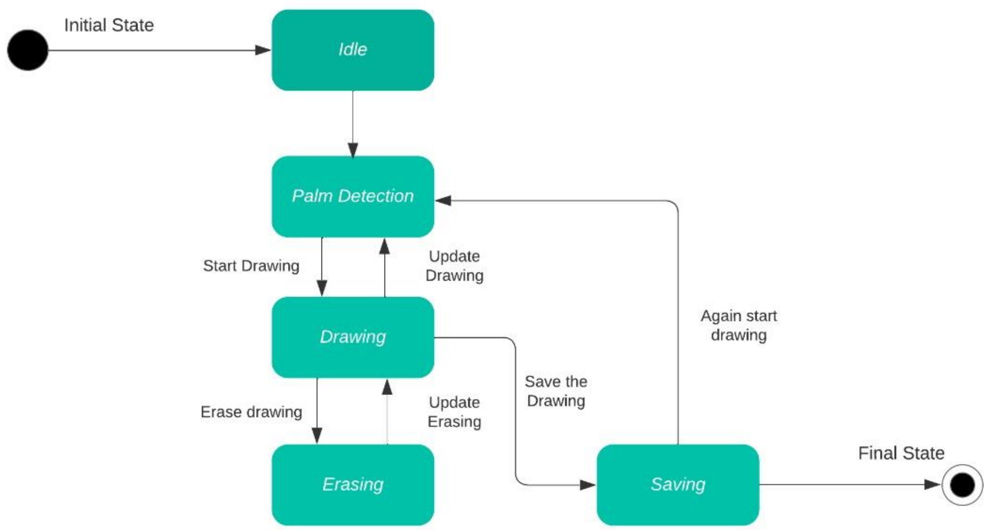
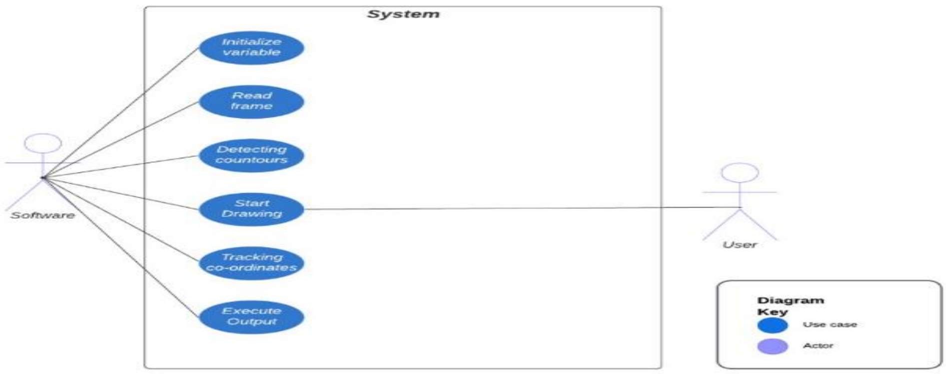


Figure 2: Use Case Diagram represents the essential interactions and functionalities of the Air Canvas system in the educational setting, emphasizing its role in facilitating effective teaching and learning experiences for students and instructors alike.



IV. IMPLEMENTATION

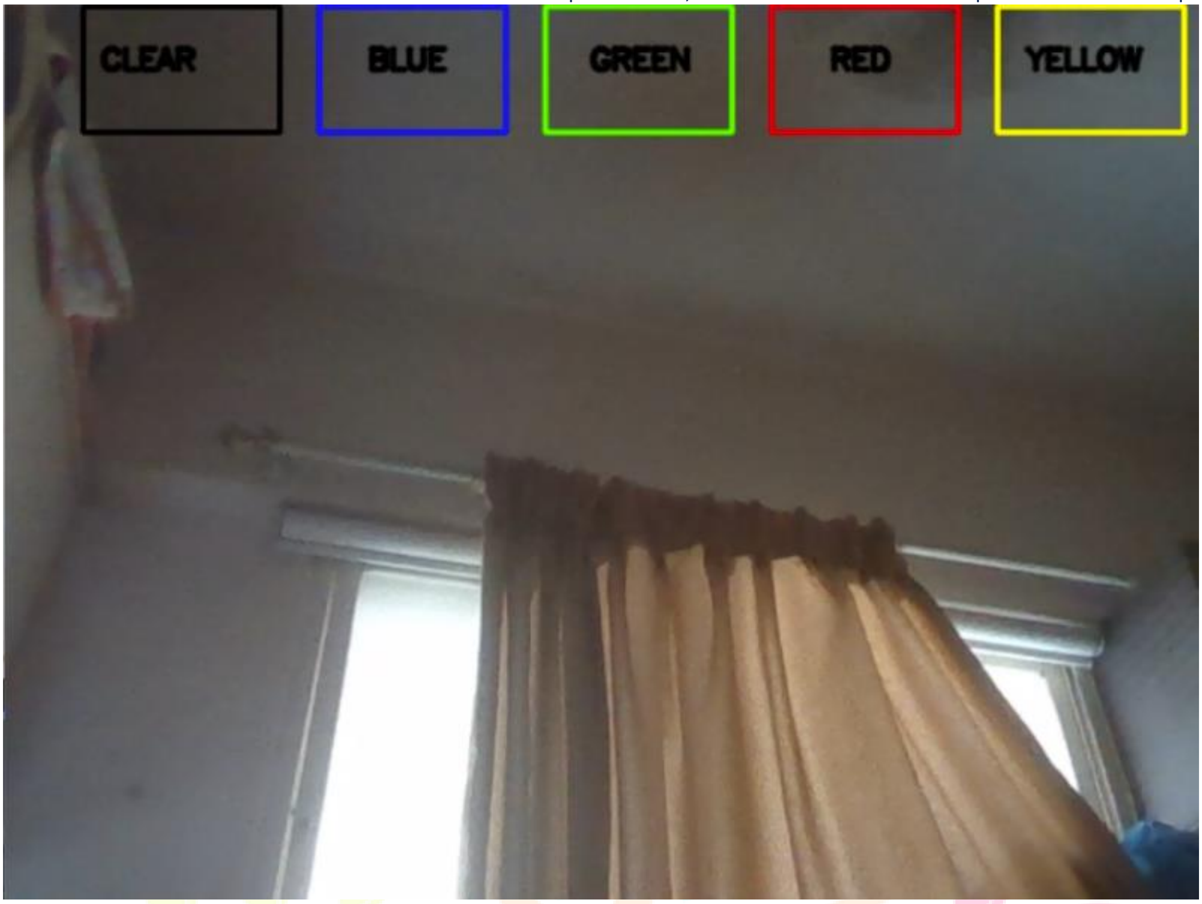
Air Canvas using Mediapipe, Opencv is the proposed system that improves the efficiency for the existing system. Each stage involved in the project and its methodologies are explained in detailed below:

1. Run the code

After installing the required libraries, executing the code triggers the automatic activation of the camera. The OpenCV frame promptly appears, adorned with an array of interactive buttons that facilitate the selection of various shapes, colours, sizes, and additional functionalities such as save, clear, and erase. This intuitive interface transforms the camera into a dynamic canvas, allowing users to seamlessly draw, modify, and manipulate visual elements using the provided tools. The inclusion of diverse options enhances the creative possibilities, making the interactive canvas an engaging and user-friendly platform for artistic expression and visual exploration.

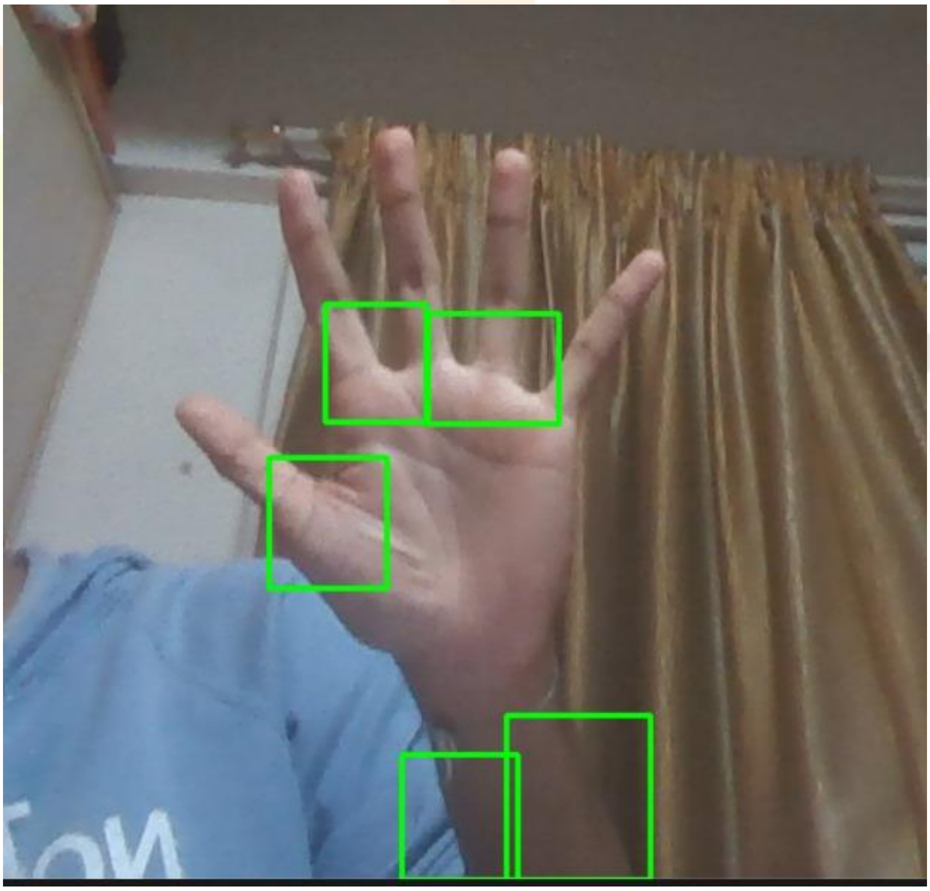
2. Camera turns on

The system initiates the webcam, commencing the recording process. Subsequently, each video frame undergoes conversion and is transmitted to the Hand Tracker class for position detection and finger tracking. Figure 2 visually represents this process, illustrating the buttons alongside the video frame, creating an interface that not only records the video in real-time but also integrates functionalities for tracking hand movements and finger positions. The simultaneous display of the recording frame and interactive buttons enhances the user experience, providing a seamless and dynamic environment for real-time video processing and hand tracking.



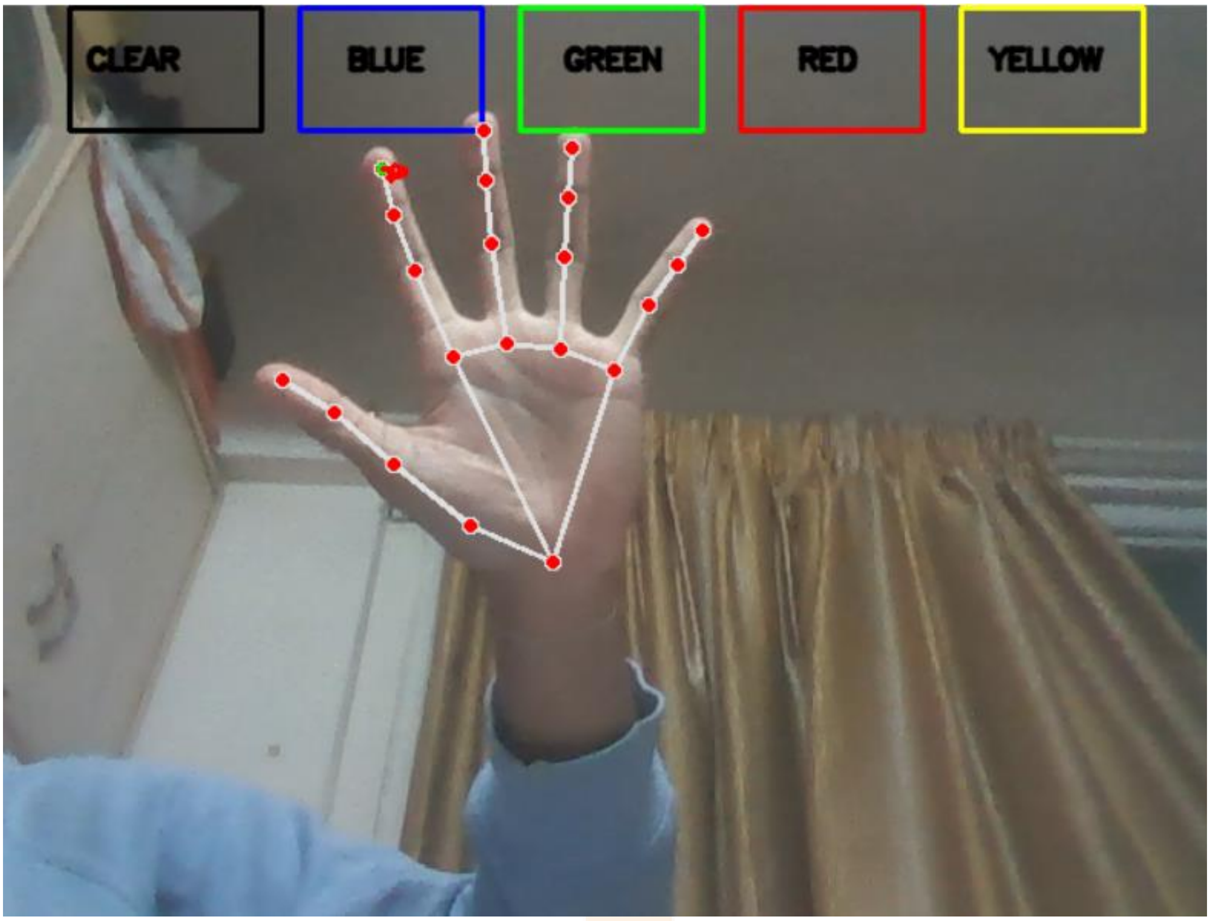
3. Tracks the hand movements

The system processes each received frame by comparing it with the landmarks provided by the Mediapipe hand tracking module. The Handtracker class plays a crucial role in this process, utilizing its "getPositions()" function to identify and extract the positions of the fingers within the frame. Additionally, the "getUpFingers()" function within the Handtracker class is employed to determine which fingers are in an open position. These two functions collectively enable the system to precisely locate and identify the positions of the fingers, offering valuable information for further processing and interaction within the application.



4. Perform actions according to buttons

The system employs the index finger's position to select various buttons. By hovering over these buttons, each with its distinct functionality, users can trigger specific actions. The interaction is intuitive, allowing users to navigate and activate different features by simply positioning their index finger over the corresponding buttons. Each button is associated with a unique function, enhancing the versatility of the system and providing a seamless and user-friendly means of executing various operations.



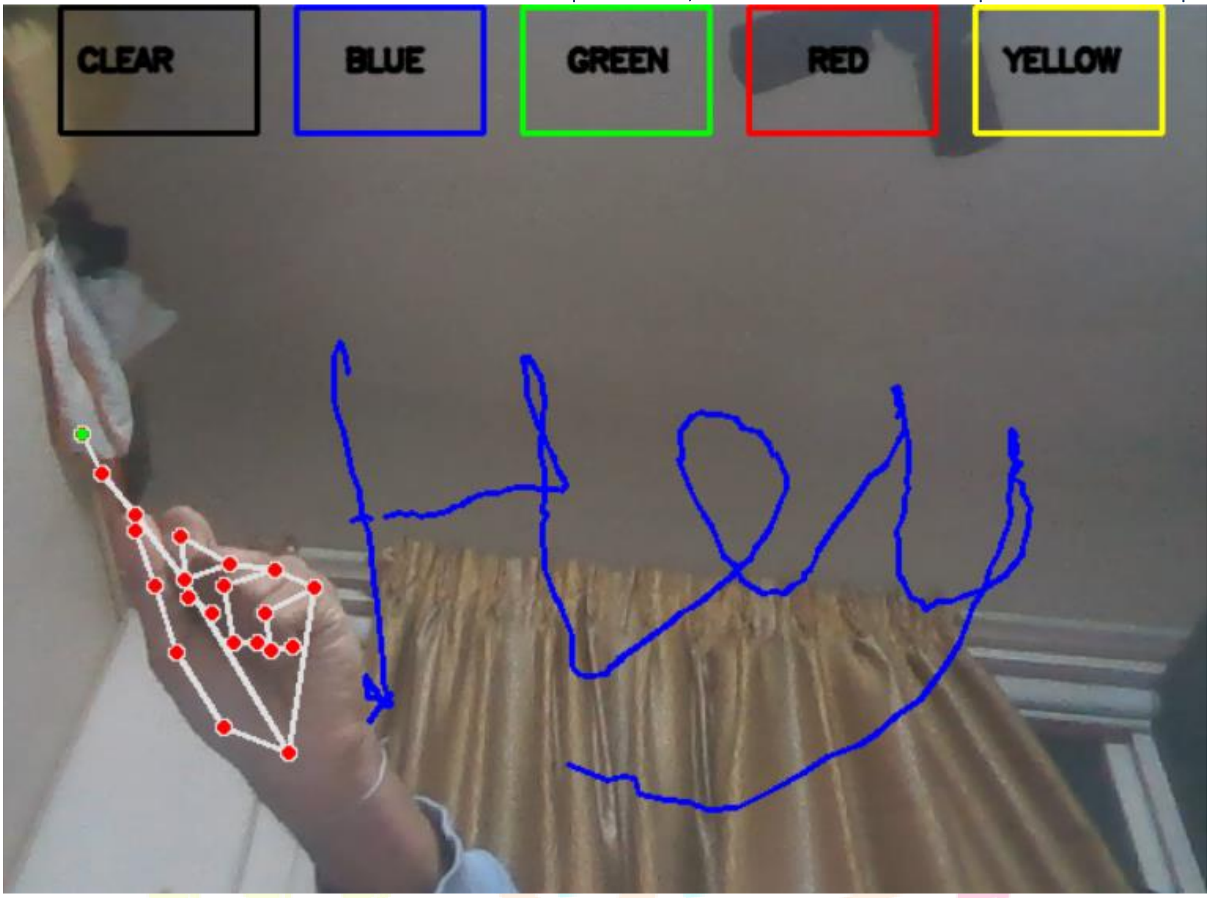
5. Depict on canvas

Here according to the hand movements and the colour selected we will be able to see the drawings on the screen and the white board.

V. RESULTS

The OpenCV library is utilized to access the camera, enabling the system to obtain a live feed. For each frame captured in real-time, the model is applied to discern the movements of the object. This dynamic tracking process unfolds concurrently with the projection of the object's movements onto the paint interface. In essence, the system synchronizes the analysis of live video frames with the real-time tracking and representation of object movements on the painting interface, creating an interactive and responsive user experience.

The below figure shows the traced text in real world:



The below figure shows the text traced on the white board:

V. REFERENCES

[1] Y. Huang, X. Liu, X. Zhang, and L. Jin, "A Pointing Gesture Based Egocentric Interaction System: Dataset, Approach, and Application," 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Las Vegas, NV, pp. 370-377, 2016.
[2] P. Ramasamy, G. Prabhu, and R. Srinivasan, "An economical air writing system is converting finger movements to text using a web camera," 2016 International Conference on Recent Trends in Information Technology (ICRTIT), Chennai, pp. 1-6, 2016.
[3] Saira Beg, M. Fahad Khan and Faisal Baig, "Text Writing in Air," Journal of Information Display Volume 14, Issue 4, 2013
[4] Alper Yilmaz, Omar Javed, Mubarak Shah, "Object Tracking: A Survey", ACM Computer Survey. Vol. 38, Issue. 4, Article 13, Pp. 1-45, 2006

[5] Yuan-Hsiang Chang, Chen-Ming Chang, "Automatic Hand-Pose Trajectory Tracking System Using Video Sequences", INTECH, pp. 132- 152, Croatia, 2010

[6] Erik B. Sudderth, Michael I. Mandel, William T. Freeman, Alan S. Willsky, "Visual Hand Tracking Using Nonparametric Belief Propagation", MIT Laboratory For Information & Decision Systems Technical Report P2603, Presented at IEEE CVPR Workshop On Generative Model-Based Vision, Pp. 1-9, 2004

[7] T. Grossman, R. Balakrishnan, G. Kurtenbach, G. Fitzmaurice, A. Khan, and B. Buxton, "Creating Principal 3D Curves with Digital Tape Drawing," Proc. Conf. Human Factors Computing Systems (CHI' 02), pp. 121- 128, 2002.

[8] T. A. C. Bragatto, G. I. S. Ruas, M. V. Lamar, "Real-time Video-Based Finger Spelling Recognition System Using Low Computational Complexity Artificial Neural Networks", IEEE ITS, pp. 393-397, 2006

[9] Yusuke Araga, Makoto Shirabayashi, Keishi Kaida, Hiroomi Hikawa, "Real Time Gesture Recognition System Using Posture Classifier and Jordan Recurrent Neural Network", IEEE World Congress on Computational Intelligence, Brisbane, Australia, 2012

[10] Ruiduo Yang, Sudeep Sarkar, "Coupled grouping and matching for sign and gesture recognition", Computer Vision and Image Understanding, Elsevier, 2008