Air Canvas using gesture control using OpenCV and Mediapipe

Chitralingappa P[1], Fardeen Hussain[2], Hemanth Kumar L[3], Dheeraj C[4], Harish Chavan M[5]

chitralingappa.cse@srit.ac.in[1], 214g1a3332@srit.ac.in[2], 224g5a3305@srit.ac.in[3], 214g1a3320@srit.ac.in[4], 214g1a3327@srit.ac.in[5]

Abstract— In recent years, one of the most intriguing and difficult study topics in the fields of image processing and pattern recognition has been writing in the air. In many applications, it can enhance the human-machine interaction and makes a significant contribution to the development of automation processes. Numerous studies have been concentrating on novel approaches and strategies that will shorten processing times while increasing recognition accuracy.

In the realm of computer vision, object tracking is regarded as a crucial problem. Object tracking techniques have gained popularity because to the development of faster computers, the availability of high-quality, reasonably priced video cameras, and the need for automated video analysis. Typically, a video analysis process consists of three main steps: first, identifying the object, then following its movement from frame to frame, and last, analyzing the object's behavior. Four distinct difficulties are considered for object tracking: object recognition, object tracking, feature selection for tracking, and choosing an appropriate object representation. Algorithms for object tracking are used in a variety of real-world applications, including autonomous surveillance, video indexing, and vehicle navigation.

The project capitalizes on this gap by concentrating on creating a motion-to-text converter that may be used as software for smart wearables that allow for writing from the air. This project reports on infrequent gestures. It will follow the finger's path using computer vision. Additionally, the created text can be utilized for a number of things, like sending emails and messages. For the deaf, it will be an effective communication tool. By doing away with the need to write, it is a productive communication tool that lowers the utilization of laptops and mobile devices.

Keywords: Air Writing, Motion-to-Text Conversion, Computer Vision, Object Tracking, Human-Machine Interaction

I. INTRODUCTION

Digital art is taking the role of traditional literature in the age of the internet. Art forms that are expressed and transmitted through digital means are referred to as digital art. The unique feature of the digital form is its reliance on contemporary science and technology. The term "traditional art" describes the kind of art that existed before digital art. Visual art, audio art, audio-visual art, and audio-visual imaginative art—which encompasses literature, painting, sculpture, architecture, music, dance, theater, and other artistic creations—can be easily separated from the recipient for analysis. Traditional and digital art are interconnected and related. Human needs are the primary motivator for social growth, even though it is not a result of popular will. In art, the same thing occurs. Given the current situation, which includes a symbiotic relationship between digital and traditional art, we must methodically comprehend the fundamentals of the relationship between the two forms of art. Pen and paper and chalk and board are two examples of the old writing methods. Developing a hand motion recognition system for digital writing is the primary goal of digital art. Writing in digital art can be done in a variety of methods, such as with a keyboard, touch screen, digital pen, stylus, electronic hand gloves, etc. However, this system uses a machine learning method to recognize hand gestures using Python programming, which fosters a natural relationship between humans and machines. The need for natural "human – computer interaction (HCI)" [10] technologies to replace conventional systems is growing quickly due to technological advancements. The rest of this paper is divided into the following categories: The additional works of literature that we consulted prior to beginning this study are included in Section 2. The difficulties we encountered in creating this system are detailed in Section 3. The problem statement that we were attempting to solve is defined in Section 4. The workflow and system methods that we used are described in Section 5. Fingertip Recognition Dataset Creation and Fingertip Recognition Model Training are two of the subsections in section 5. Workflow algorithm, Section 6.

II. LITERATURE REVIEW

The suggested approach in [3] detected the hand shape by using the color and depth data from the Kinect sensor. Despite the Kinect sensor, gesture recognition is still possible. It remains an extremely difficult topic. This Kinect sensor has a resolution of just 640 x 480. It performs admirably when tracking a big object, like the human body. However, it is difficult to follow a little object like a finger.

The technique proposed by the authors in [4] involves mounting an LED on the user's finger and using the webcam to track the finger. The character in the database is compared to the one that was drawn. The alphabet that corresponds to the drawn pattern is returned. It needs an LED directed light source that is red and fastened to the finger. Additionally, it is believed that the webcam's focus is empty of anything red other than the LED light.

The use of an augmented segmented desk interface for interaction was suggested in [5]. This technology uses a charge-coupled device (CCD) camera and a video projector so that users can control desktop apps with their fingers. Each hand in this system has a distinct function. While the right hand is used to choose manipulable items, the left hand is used to select radial menus. It makes use of an infrared camera to accomplish this. Because it is computationally costly to determine the fingertip, this method defines search windows for fingertips.

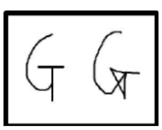
III. CHALLENGES IDENTIFIED

A. Fingertip recognition

There are no highlighters, paints, or family members in the current system, which simply requires your fingers to operate. Without a depth sensor, it can be very difficult to recognize and describe an object, such a finger, from an RGB image.

B. Insufficient pen-up and pen-down movements

The system writes from above using a single RGB camera. Pen motions up and down cannot be followed since depth sensing is not feasible. As a result, the whole trajectory of the fingertip is tracked; the final image would be ridiculous and unrecognizable to the model. Figure 1 illustrates the distinction between handwriting and airwriting the letter "G."



C. Managing the system in real time

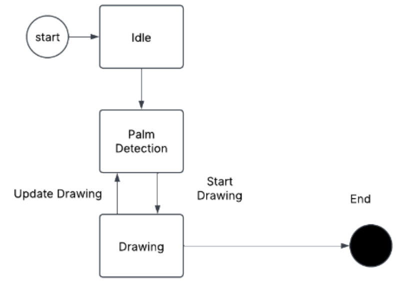
It takes a lot of programming attention to switch the system between states using real-time hand gestures. To properly control his plan, the user also needs to be familiar with a variety of movements.

D. Limited features and gestures

The tools built for computer vision can be very limited as it takes some knowledge to understand the CV techniques, we can use multiple gestures and tools for making the canvas more versatile.

IV. PROPOSED SYSTEM

A system design for a "Air Canvas" application using OpenCV needs several key components. This app enables users to utilize camera-captured motions to sketch or paint in midair. Below is a simplified system architecture with a flowchart.



User Interface (UI): This is the front-end component that interacts with the user. There are options for starting, ending, and personalizing the drawing session. It displays the drawing canvas for users to utilize and provides options for selecting colors and shapes in addition to cleaning the board.

Camera Module: This module records GDG frames in real time by using the computer's camera as an interface and it also provides the video stream for the Image Processing Module. The module absorbs the video from the camera stream and recognizes human gestures by processing and analyzing images in real time.

V. ARCHITECTURE

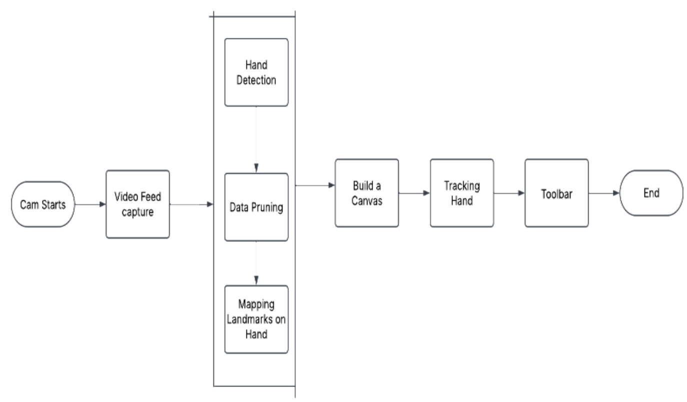
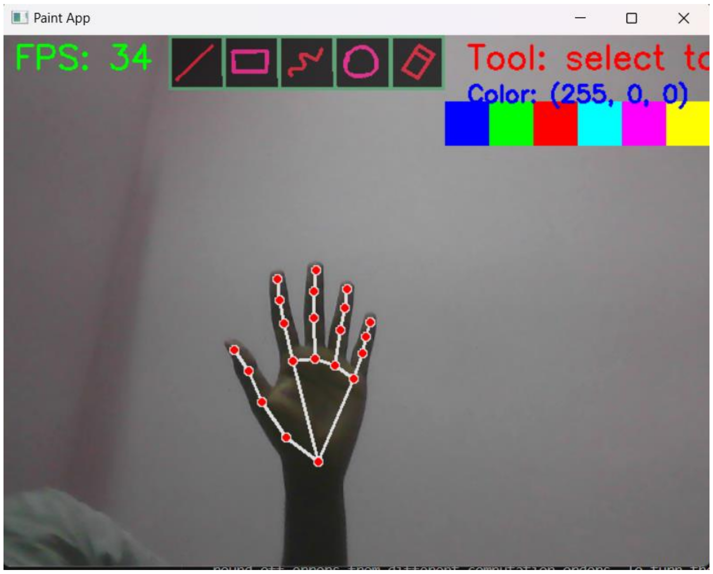


Fig.1 Architecture of the project

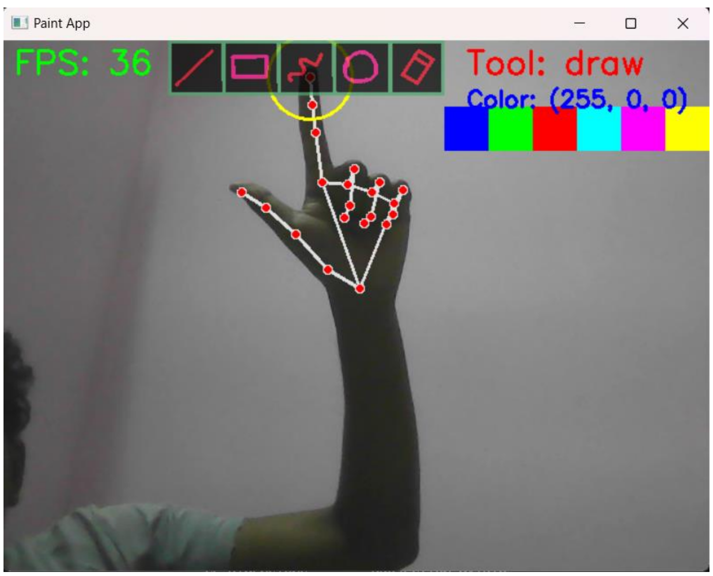
The flowchart illustrates the process of air writing recognition using computer vision. It begins with the camera activation, followed by video feed capture. The next stage involves hand detection, where raw data is refined through data pruning and landmark mapping on the hand to improve tracking accuracy. Once the essential hand landmarks are identified, a virtual canvas is built to facilitate writing. The system then tracks the hand movements, ensuring precise recognition of gestures. A toolbar is integrated to allow interaction, such as selecting colors or clearing the canvas. Finally, the process concludes, making the system efficient for air writing applications.

VI. RESULTS AND DISCUSSIONS

This is the first interface that has the toolbar and fps monitor; we can get it by executing the Python script that uses the OpenCV and mediapipe modules.



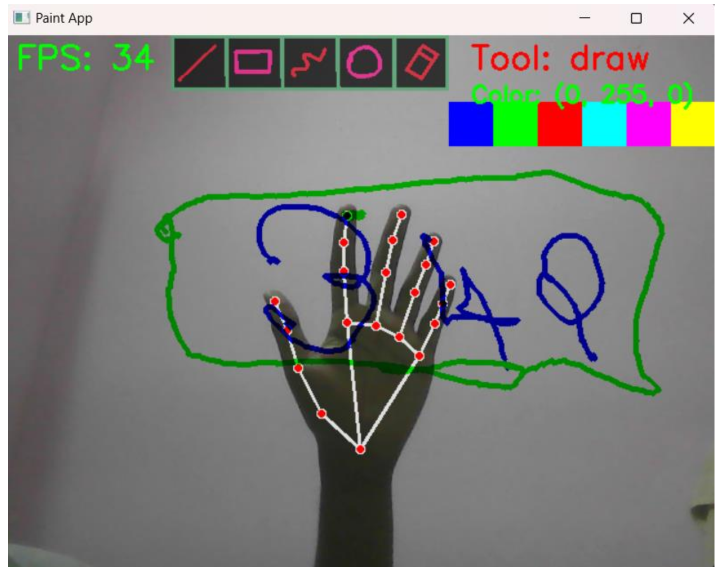
To write and draw, we can choose from five distinct tools and six different colors. Addittionaly we have an option for erasing the lines using eraser tool.



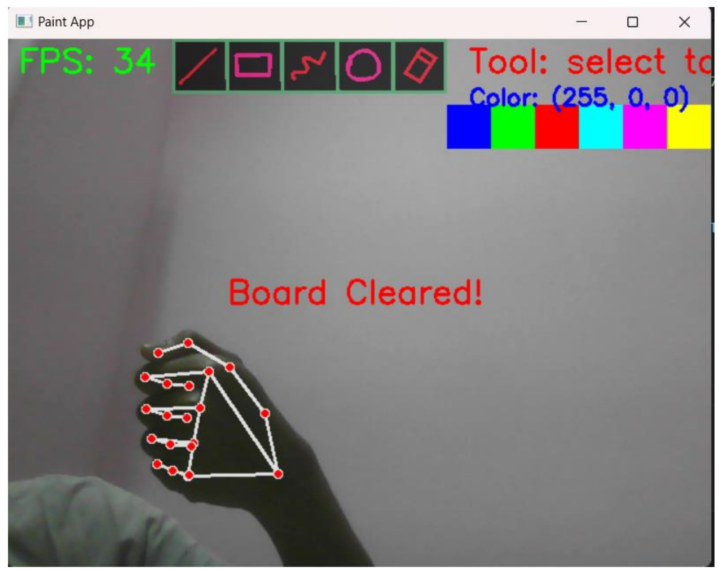
The most dependable tool for writing and drawing is the draw tool. When choosing a tool, we must close the second finger in order to isolate the tip. We can draw after opening the second finger, and so on.

We can use different shapes using line, rectangle and circle tools for adding more versatility to our canvas, the fps monitor is also useful to depict the improvements needed in the hardware the user has currently.

The right side of the canvas shows us the current tool selected and also the color.



The paths the index finger takes when drawing are depicted in the figure above.



We may use the following fist motion to clear the board, after clearing the stuff we have drawn thus far, the message "Board Cleared" appears on the screen.

VII. FUTURE ENHANCEMENTS

Future enhancements of the air canvas system will focus on improving accuracy, responsiveness, and usability. Advanced machine learning models can be integrated to enhance gesture recognition, reducing errors and improving adaptability to different hand movements. Adding multi-finger gesture support could introduce new functionalities, such as zooming, rotating, or selecting specific tools. Augmented reality (AR) integration may provide real-time visual feedback, making the writing experience more immersive. Furthermore, cloud-based processing could allow real-time synchronization across multiple devices, enabling collaborative writing or remote communication. Enhancements in user interface design, such as customizable toolbars and gesture-based shortcuts, will also improve accessibility and user experience.

VIII. CONCLUSION

To sum up, creating a "Air Canvas" application with OpenCV provides an engaging and useful means of investigating the potential of computer vision technologies. This application gives users a creative way to create digital art by allowing them to draw on a virtual canvas using their hand motions. A number of important topics can be emphasized throughout the development process challenges.

IX. 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 P 2603, 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

[11] R. Wang, S. Paris, and J. Popovic, "6D hands: markerless hand-tracking for computer-aided design," in Proc. 24th Ann. ACM Symp. User Interface Softw. Technol., 2011, pp. 549–558.

[12] Maryam Khosravi Nahouji, "2D Finger Motion Tracking, Implementation For Android Based Smartphones", Master's Thesis, CHALMERS Applied Information Technology,2012, pp 1-48

[13] EshedOhn-Bar, Mohan ManubhaiTrivedi, "Hand Gesture Recognition In Real-Time For Automotive Interfaces," IEEE Transactions on Intelligent Transportation Systems, VOL. 15, NO. 6, December 2014, pp 2368-2377

[14] Kenji Oka, Yoichi Sato, and Hideki Koike, "Real-Time Fingertip Tracking and Gesture Recognition," IEEE Computer Graphics and Applications, 2002, pp.64-71.

[15] H.M. Cooper, "Sign Language Recognition: Generalising to More Complex Corpora", Ph.D. Thesis, Centre for Vision, Speech and Signal Processing Faculty of Engineering and Physical Sciences, University of Surrey, UK, 2012.

[16] Vladimir I. Pavlovic, Rajeev Sharma, and Thomas S. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review," IEEE Transactions on Pattern Analysis and Machine Intelligence, VOL. 19, NO. 7, JULY 1997, pp.677-695

[17] Guo-Zhen Wang, Yi-Pai Huang, Tian-Sheeran Chang, and Tsu-Han Chen, "Bare Finger 3D Air-Touch System Using an Embedded Optical Sensor Array for Mobile Displays", Journal Of Display Technology, VOL. 10, NO. 1, JANUARY 2014, pp.13-18

[18] Napa Sae-Bae, Kowsar Ahmed, Katherine Isbister, NasirMemon, "Biometric-rich gestures: a novel approach to authentication on multi-touch devices," Proc. SIGCHI Conference on Human Factors in Computing System,2005, pp.977-986

[19] W. Makela, "Working 3D Meshes and Particles with Finger Tips, towards an Immersive Artists' Interface," Proc. IEEE Virtual Reality Workshop, pp. 77-80, 2005.

[20] A.D. Gregory, S.A. Ehmann, and M.C. Lin, "inTouch: Interactive Multiresolution Modeling and 3D Painting with a Haptic Interface," Proc. IEEE Virtual Reality (VR' 02), pp. 45-52, 2000.

[21] W. C. Westerman, H. Lamiraux, and M. E. Dreisbach, “Swipe gestures for touch screen keyboards,” Nov. 15 2011, US Patent 8,059,101

[22] S. Vikram, L. Li, and S. Russell, "Handwriting and gestures in the air, recognizing on the fly," in Proceedings of the CHI, vol. 13, 2013, pp. 1179–1184.