

# Air Canvas: A Gesture-Based Platform for Interactive Digital Drawing and Real-Time Creativity

Radhika Soni<sup>1</sup>, Priyanka Sharma<sup>1</sup>, Payal Nagar<sup>1</sup>, Ajay Bhardwaj<sup>2</sup>, Nikhar Bhatnagar<sup>3</sup>

<sup>1</sup>Department of Computer Science & Engineering, Swami Keshvanand Institute of Technology, Management & Gramothan, Jaipur, Rajasthan-302017 (India)

<sup>2</sup>Department of Electrical Engineering, Swami Keshvanand Institute of Technology, Management & Gramothan, Jaipur, Rajasthan-302017 (India)

<sup>3</sup>Department of Information Technology, Swami Keshvanand Institute of Technology, Management & Gramothan, Jaipur, Rajasthan-302017 (India)

*Email:* radhikasoni.jpr@gmail.com, priyanka.sharma@skit.ac.in, payalnagar317@gmail.com, ajay.bhardwaj@skit.ac.in, nikhar.bhatnagar@skit.ac.in

Received 09.05.2025 received in revised form 19.09.2025, accepted 19.09.2025

DOI: 10.47904/IJSKIT.15.2.2025.19-23

**Abstract-** In a world where technology evolves faster than ever, finding ways to interact with it naturally is more important than ever. If we talk about “drawing” it’s very significant because it is the most basic and powerful form of communication. It’s a personal and universal tool which connects thinking and feeling. The transition from drawing on walls of ancient caves to creating digital art on smartphones is an example of the remarkable journey of humans in innovation and technology. This technology gave us various options for drawing. “Air Canvas” helps to portray the idea of drawing in air. This study introduces a gesture-based system that lets people draw digitally in real time—using just their hands and a regular webcam. The primary objective of this study is to enable users to draw on a digital screen without any physical contact, simply by moving a colored object (such as a fingertip covered with a blue cap) in front of a webcam providing an easy, accessible and touchless way of engaging with digital spaces. The uniqueness of this study lies in using technologies like OpenCV and deep learning and building features like recognizing hand or object gestures accurately, enabling them to make refined shapes and designs. A lot of practical applications can be supported through this - in domains like education - digital boards, accessibility-focused tools like screen readers, screen magnifiers, art and games. It can increase collaboration in online education and meeting systems where people can convey their ideas and points on a shared canvas which everyone can see. There’s a lot which this application provides like allowing users to change tools, change ink colours, changing background colour, undo-redo and to save their work and all these can be accessed through simple hand movement, making it more effortless and engaging. Overall, it will help to span the distance between user and virtual technology. This study emphasizes the power of gesture recognition technology in transforming the way we engage with computers.

**Keywords-** Air Canvas, finger-tip, OpenCV, Hand Tracking

## 1. 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 [1]. This platform leverages the power of hand gesture recognition, computer vision, and Python programming to create an interactive and immersive learning experience for students. The Air Canvas Study is an innovative application that combines computer vision and gesture recognition to enable users to draw and write in the air using their fingers. By leveraging the power of OpenCV and NumPy libraries, this study revolutionizes the way we interact with digital interfaces and features a motion-to-text converter [2].

This study is built using the computer vision techniques of OpenCV which is an open-source library. Although Python is the preferred language due to its extensive libraries and easy-to-use syntax, any OpenCV-compatible language can be used, provided you understand the basics. To achieve the objective, Color Detection and Tracking techniques are also used. The color marker is detected and a mask is created. As part of the process, further morphological operations are performed on the mask-- where the mask that has been eroded reduces the impurities present, and a mask that has been dilated further restores the mask that has been eroded [3]. This study allows users to draw and create art in a virtual canvas using their hand gestures. It combines computer vision techniques (OpenCV and media pipe library) to recognize hand gestures and detect objects on the canvas. The way we engage with digital technology is rapidly evolving [4]. More and more, people are looking for interaction methods that feel instinctive—ones that don’t require constant

tapping or typing Computer vision study implemented with OpenCV with Machine learning using the Media pipe.

The use of OpenCV's computer vision techniques is what we will be using to construct this study. Python is kept as the preferred language because of its libraries which align with the need of this study [5]; however, the fundamentals can be applied in any language that supports OpenCV.

The Air Canvas Study is an innovative application that combines computer vision and gesture recognition to enable users to draw and write in the air using their fingers.

The study features a motion- to-text converter. 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.

## 2. LITERATURE REVIEW

Gesture-based interaction is altering the way we connect with digital systems, presenting a more streamlined and comprehensible alternative to traditional tools. Hand movements, unlike keyboards, touchscreens, or styluses, can create a more natural experience, unlike those that can sometimes feel limiting. Education, art, and virtual teamwork are areas where freedom of movement can spark creativity and foster collaboration, making this approach particularly valuable [6]. Although gesture recognition is gaining momentum, numerous current systems are still struggling to accurately interpret natural movements. In the end, systems that are intuitive and contact-free can transform the way we experience the digital world, making it more engaging and accessible to us.

The recognition model they developed 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 Intel Sense 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.

Movement recognition using LED: The system has a camera to detect the movement of the LED mounted on

the fingertip. It can recognize the pattern drawn on the screen by the LED. Although it has advantages like fast movement tracking and accurate results, the disadvantage is that the LED must be red only and not any other color [7].

Hand Recognition with Kinect Sensor: The Kinect sensor is utilized to obtain depth and color data for the purpose of detecting hand movements. However, gesture recognition remains a complex challenge. While the Kinect sensor operates effectively with larger objects, it may struggle to function properly with smaller items [8]. The use of automatic object tracking is extensive in computers, including computer vision and human machine interaction. There are various suggestions for tracking algorithms in the literature.

One team of researchers used it to interpret the signals languages, some to see hand gestures, another text-tracking group as well recognition, and body monitoring the visible movement of objects as well character recognition based on finger tracking, etc. Bragatto et al. he built a way for that automatically translates Brazilian Sign Language from video input. They used NN in two stages: color recognition and steps to go check hand shape. Their results show that the method proposed well works with the acquisition rate of 99.2%. Cooper also introduced the method managing the most complex 3D cell bioprinting there is a standard set. Cooper developed a process that minimizes tracking by identifying errors in the division procedures and tracking. Cooper used two treatment modalities; One is for his movement, and the other is used for clarification hand shape. He used the screen for expand is vocabulary [9]-[13]. Viseme is an important position word of mouth in pronouncing the word a Phoneme and visual representations of phonemes. Over time, you become less formal a way to identify the characters.

## 3. PROPOSED SYSTEM

The proposed framework includes cameras and screens that will be used to read inputs and display outputs. The shapes we need to draw on the output screen are being drawn using our fingers. We need to ensure that the camera can properly detect our hand, in order the movements of our fingertips can be smoothly detected and recognized. Hand tracking involves use of computer vision for hand movement. It allows us to create variety of programs that use gestures and coordinates as an input. In our program, hand tracking is performed using the same code across different studies. This problem can be solved by creating a hand tracking module because we only need to write the code once. Converting this code into a module is our next step:

Gesture-based interaction is altering the way we connect with digital systems, presenting a more streamlined and comprehensible alternative to traditional tools. Hand movements, unlike keyboards, touchscreens, or styluses, can create a more natural experience, unlike those that can sometimes feel limiting. Education, art, and virtual teamwork are areas where freedom of movement can spark creativity and foster collaboration, making this approach particularly valuable [6]. Although gesture recognition is gaining momentum, numerous current systems are still struggling to accurately interpret natural movements. In the end, systems that are intuitive and contact-free can transform the way we experience the digital world, making it more engaging and accessible to us.

The recognition model they developed 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 Intel Sense SR300 camera in conjunction with a computer.

A cropped image of the hand is provided by Media Pipe after processing the complete input image for palm detection. Media Pipe spotted 21 hand markers in the hand image that was cropped. Video tracking involves an algorithm that analyzes sequential video frames and outputs the movement of targets between frames. There are various algorithms apart from this, each with its own pros and cons. It's crucial to consider the intended use while picking up the appropriate algorithm. Target representation and localization are the main components of a visual tracking system, along with filtering and data association.

Video tracking involves capturing a transitioning object over intervals.

Several use cases for it can be found in technologies like traffic control, security and surveillance, video communication and compression, AR, medical imaging, and video editing are among the many uses it has. The system's adaptability allows it to accommodate the diverse needs of educators, remote teams, creative professionals and artists.

In this proposed framework, we are going to utilize a webcam and display it on a monitor screen. Here, we will be using a pen or hand to draw attractive images in front of the camera. Then, we will attract those images and show them on the presentation unit. Our framework is suitable for decoding time-collection pace boom alerts into extensive vectors. The pen can be utilized by users to compose numbers or make hand motions, and this can be demonstrated at the presentation unit.

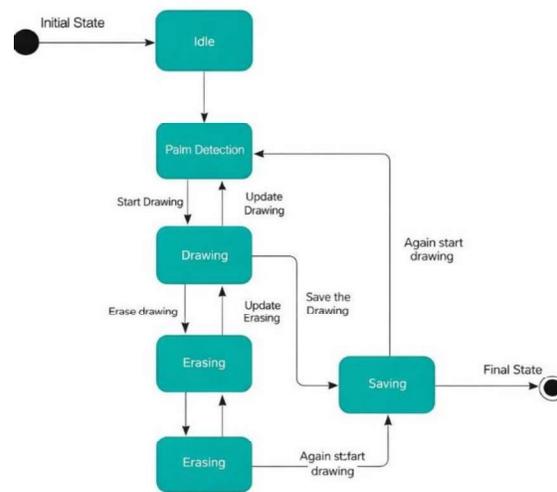


Figure 1: Workflow

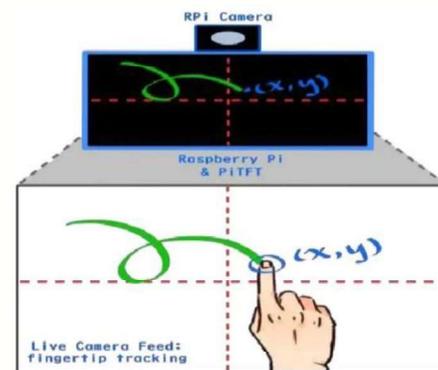


Figure 2: Picture representation of Air Canvas

#### 4. BLOCK DIAGRAM

**Camera:** It uses optical technology to capture images that can be stored directly, transmitted to another location, or both. The term camera is derived from the Latin word 'Camera Obscura', which means 'dark chamber'. The modern camera is a descendant of the obscura camera, and its functions are very similar to those of the human eye. It is required to capture movements. There is no need of any branded camera, it should just clearly concentrate on finger movements. Also, mind appropriate gap between camera and you. Cameras are necessary hardware for the system to capture your hand movements.

**Screen:** The hand actions and output will be shown on the screen. On the screen, the camera will show whatever view it is recording. By viewing on the screen, you can easily perform your drawing action. It is crucial that your output is displayed on the screen. Python is a programming language that is widely used for general-purpose and high-level tasks. The Python Software Foundation was responsible for developing it after Guido van Rossum created it in 1991. Code readability was the

main focus of its design, and it allowed programmers and developers to convey their ideas in lesser lines.

**Media Pipe:** Media Pipe is the Google Framework which helps solve the problems in hand tracking. Media Pipe has ready-to-use machine learning solutions that can be used in various machine learning projects. Also, it contains other modules like movement recognition, gestures recognition, and some others. Enhancing user experience across various technological domains and platforms can be achieved through detecting the movement of hands. This capability serves as a basis for hand gesture recognition and understanding sign language.

Furthermore, it enables the overlay of digital information and content onto the real world in augmented reality. Despite being a natural skill for humans, effective real-time hand perception presents a significant challenge in computer vision, as hands often obscure one another and do not exhibit high-contrast patterns. Media Pipe Hands offers a sophisticated hand and finger tracking solution that utilizes machine learning to identify 21 three-dimensional hand structures from each single image. This technology is capable of tracking multiple hands simultaneously, in contrast to current leading systems that predominantly depend on robust desktop or laptop environments for processing. By building this advanced hand perception capability accessible to a broader research and development community, we will foster the emergence of novel applications and stimulate new avenues of research.

**OpenCV:** OpenCV is a library which is used for image recognition. It will identify our hand tracking and drawing. It is a library basically designed to work on image processing and image recognition. The OpenCV computer vision library encompasses various methods for object detection and image processing. By leveraging this library within the Python programming language, developers can create applications that operate in real-time. The OpenCV library is instrumental in processing both videos and images, along with implementing complex techniques such as object and face detection.

A zero-based index allows for the retrieval of elements within the collection. Each item in a multidimensional (ND) array occupies an equal proportion of space within a memory block. Each element in the ND array is represented as a data-type object known as 'type'. When an item is accessed from an ND array object through slicing.

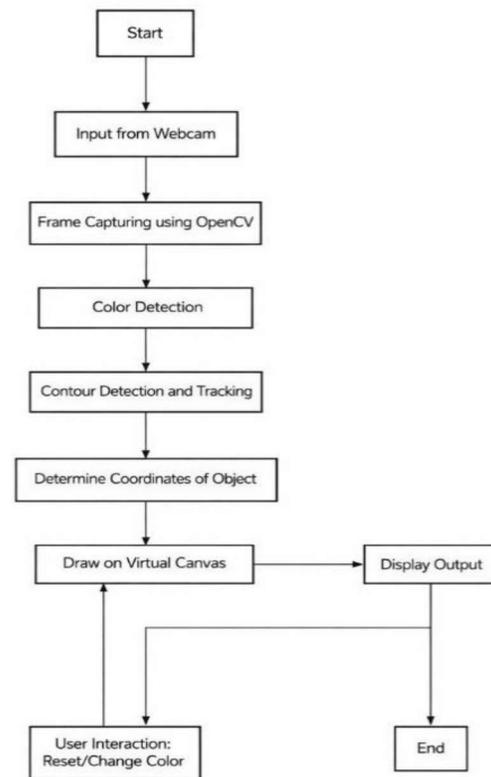


Figure 3: Flow Chart

## 5. RESULTS

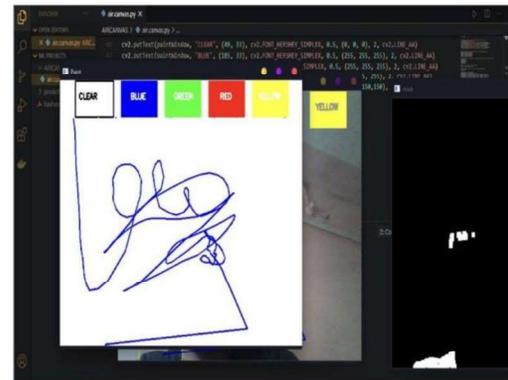


Figure 4: Real time drawing output using hand gestures

## 6. CONCLUSION

Air Canvas is a fun and interactive application that demonstrates the capabilities of OpenCV for object detection and tracking, and for drawing on images. It can be used as a creative tool for digital art, or as a learning project for students interested in computer vision and image processing. Power of Python, Media Pipe and OpenCV have been used to develop this project which will allow users to create drawings in air using their

fingertips. Options like adding various colors, eraser or clearing the whole screen, and undo redo options are available. Future scope of this application can be for virtual teaching purpose, can eliminate the need to use keyboard especially helping for those who can't see or find difficulty in using the keyboard and helping in communication. The scope can be expanded by integrating IoT devices to build more great products for various industries in designing.

## REFERENCES

- [1]. M. I. Memon, S. M. N. Iqbal, and A. W. R. Elbattah, "Real-time hand gesture recognition for human-computer interaction," *International Journal of Computer Applications*, vol. 154, pp. 12–16, (2017).
- [2]. S. Agarwal, S. K. Chaurasia, and A. Goel, "Gesture recognition with OpenCV and Python," *Proc. Int. Conf. Intelligent Computing and Control Systems*, vol. 1, pp. 97–101, (2019).
- [3]. P. Sharma, R. Singh, and H. Sharma, "Interactive digital drawing with real-time hand gesture recognition," *International Journal of Computer Vision and Image Processing*, vol. 12, pp. 56–68, (2020).
- [4]. T. S. Rajpoot and V. Gohil, "A survey on hand gesture recognition for human-computer interaction," *International Journal of Computer Science and Information Technologies*, vol. 6, pp. 1622–1632, (2015).
- [5]. J. R. Smith and R. A. Lee, "Gesture-based drawing systems for educational purposes," *Proc. IEEE Int. Conf. Education and Information Systems*, vol. 1, pp. 221–225, (2018).
- [6]. A.K. Gupta, "Interactive drawing with hand gesture recognition using OpenCV," *International Journal of Advanced Research in Computer Science and Engineering*, vol. 6, pp. 1793–1798, (2019).
- [7]. D. T. Nguyen, S. F. M. Ahmed, and C. L. Tan, "Deep learning for real-time gesture recognition in interactive applications," *IEEE Transactions on Human-Machine Systems*, vol. 48, pp. 233–240, (2018).
- [8]. P. Sharma and R. Rathi, "Creating gesture-controlled interfaces for interactive learning via computer vision," *Journal of Computer Science and Technology*, vol. 11, pp. 25–32, (2021).
- [9]. C. Zhang, L. Wang, and X. Li, "A virtual whiteboard system for gesture control using hand tracking and machine learning," *Journal of Virtual Reality and Broadcasting*, vol. 13, pp. 45–52, (2020).
- [10]. M. R. Gadekallu and N. D. R. Mahesh, "Deep learning models for hand gesture recognition in interactive digital applications," *International Journal of Computer Vision*, vol. 42, pp. 210–220, (2022).
- [11]. J. Smith and A. Gupta, "Enhancing real-time gesture recognition with multi-stream CNNs," *Journal of Human-Computer Interaction Research*, vol. 45, pp. 98–115, (2023).
- [12]. L. Zhao, H. Wang, and Y. Chen, "Gesture recognition frameworks for augmented reality applications," *Journal of Virtual Reality Interfaces*, vol. 17, pp. 25–37, (2024).
- [13]. T. Nguyen and F. Ahmed, "Deep learning approaches to real-time hand gesture recognition," *IEEE Transactions on Human-Machine Systems*, vol. 52, pp. 345–356, (2023).
- [14]. K. Tanaka and S. Mori, "Contactless interaction using gesture-based systems: A review," *International Journal of Computer Vision*, vol. 76, pp. 420–430, (2023).
- [15]. R. Patel and P. Iyer, "Cost-effective gesture control for digital art platforms," *Proc. Int. Conf. Innovative Computing*, vol. 34, pp. 67–78, (2024).
- [16]. G. Gupta, P. Sharma, and A. Joshi, "Real-time hand gesture recognition and distance estimation with webcam integration," *Proc. 6th Int. Conf. Information Management & Machine Intelligence (ICIMMI)*, vol. 1, pp. 1–5, (2025).