Virtual Air Canvas

Akhil Mohan
Student, Dept of Computer Science and Engineering
Universal Engineering College
Thrissur, Kerala
akhilmohan8547@gmail.com

Sreema E R
Assistant Professor, Dept of CSE
Universal Engineering College
Thrissur, Kerala
sreema@uec.ac.in

Leshma Mohandas
Student, Dept of Computer Science and Engineering
Universal Engineering College
Thrissur, Kerala
leshmamohandas2002@gmail.com

Prabath P U
Student, Dept of Computer Science and Engineering
Universal Engineering College
Thrissur, Kerala
prabathunni826@gmail.com

Saeedh Mohammed
Student, Dept of Computer Science and Engineering
Universal Engineering College
Thrissur, Kerala
saeedhmohd244@gmail.com

***Abstract*—An important development in human-computer interaction is this creative use of open-source frameworks to create gesture-controlled whiteboard technology. The device allows users to interact with a virtual whiteboard using natural gestures by utilizing advanced computer vision algorithms for hand tracking and interpretation. This strategy enhances relationships and interactions by fostering new forms of communication in addition to expanding upon conventional writing techniques. The system improves accessibility and provides opportunities for people with a range of communication challenges by utilizing contemporary technology. Its ability to improve people's quality of life by providing new channels for communication and giving them new tools for expression and involvement is highlighted by its contribution to the changing field of human-computer interaction.**

***Index Terms*—Data Processing, Handmark Estimation, Computer Vision algorithms, MediaPipe, OpenCV.**

I. Introduction

Writing in the air has been the subject of fascinating and compelling study in the fields of image processing and pattern recognition in recent times. It can enhance the human-machine interface in a variety of operations and significantly advance the automation process. Several studies have concentrated on cutting-edge methods and configurations that provide improved recognition precision at reduced processing times.

One of the system's most notable features is its virtual whiteboard, which raises the bar for the precision and legibility of lines in digital airwriting. It achieves this by fusing modern fonts with hand gestures. With the ability to use their hands as a virtual paintbrush and a fluid, organic medium for artistic expression, this virtual canvas transforms how people interact with digital art. Users can redefine the creative process by expressing themselves through writing, painting, and other media using their natural hand movements.

Virtual Air Canvas is unique because of its accuracy. In addition to identifying motions, it records the exact movement of your fingertips, allowing you to create delicate and complex digital artwork. It can monitor finger movements and interpret hand gestures with amazing accuracy by utilizing cuttingedge technologies like MediaPipe and OpenCV, giving the experience a very realistic feel.

The system's extensive tracking of fingertip motions allows users to construct their digital masterpieces with precision and subtlety. This degree of detail opens up new avenues for precise control and manipulation in the virtual world, while also broadening the platform's creative possibilities. Powered by cutting-edge technology, Virtual Air Canvas uses MediaPipe and OpenCV. Thanks to OpenCV, a well-liked computer vision framework, the system can understand and respond to the user's hand movements with unexpected accuracy. Furthermore, MediaPipe technology facilitates finger tracking, ensuring a seamless and natural extension of the user's intentions onto the digital canvas.

With the system's use of Optical Character Recognition (OCR) technology, character recognition in the context of digital art creation has experienced a revolutionary change. This innovative improvement raises the bar for character recognition's accuracy and dependability to unprecedented levels. Through the use of OCR, the technology can precisely distinguish individual letters from continuous air-written movements, providing users with a more complex and responsive experience. OCR not only enhances the technical aspects of character recognition but also provides users with additional creative alternatives. Strong connections between written and visual expression can be made in digital artwork by skillfully fusing textual and graphic elements. The story can now be expanded by users by simply adding text, messages, or intricate embellishments to their artistic creations.

The system, which places great emphasis on innovation, redefines the creative process in digital art through the deft integration of computer vision, gesture recognition, and OCR technologies. This means that on an engaging and user-friendly platform, users can build visually stunning compositions using only their hand gestures. With its modern typefaces, clean lines, and markerless virtual whiteboard capabilities, this system is a trailblazer that has elevated the standard for interactive and expressive digital art experiences. The landscape of virtual artistic expression is always evolving, and this system sticks out as a trailblazing force that is expanding the bounds of what is feasible at the intersection of technological innovation and artistic genius.

II. Theory and Related Works

Taiki Watanabe and his team [1] are studying air-writing as a kind of human-computer interaction, which has recently gained popularity. Air-writing relies on gesture recognition to connect natural language with digital interfaces. The researchers suggest powerful character recognition methods to address variances in writing styles and articulation speeds. The authors propose a novel solution using a webcam-based air-writing system, as other approaches using gyroscopes, accelerometers, or computer vision have limitations. They propose a hybrid deep learning model that combines BiLSTM models for sequential data processing and CNNs for image recognition.The goal of the effort is to overcome the current lack of comprehensive and accurate air-writing recognition systems, enhance human-computer interaction paradigms, and enhance air-writing recognition. It highlights the importance of freely accessible datasets, like the 6DMG dataset.

Saez-Mingorance, Mendez-Gomez, Mauro, Castillo Morales, Pegalajar Cuellar, and Morales-Santos [2] explore the evolution of human-computer interaction and the use of natural communication interfaces. They focus on air-writing, a compelling application that allows users to express characters through hand movements in open space. This study utilizes ultrasonic transceivers to collect hand movement trajectories for air-writing recognition, offering a more efficient alternative compared to traditional methods. The main contribution is a suggestion for an air-writing instrument with ultrasonic technology to enhance character identification accuracy. The authors evaluate convolutional LSTM as the most successful deep learning algorithm, surpassing CNN, LSTM, convolutional autoencoder, and convolutional LSTM. With a phenomenal accuracy rate of 99.51%, it is the clear winner. Validation in real-world settings evaluates the system's utility and effectiveness in data collection and classification. The employment of ultrasonic transceivers and deep learning algorithms has greatly enhanced air-writing recognition systems' reliability and effectiveness.

Chen. H and their team's [3] innovative acoustic wave airwriting approach enhances human-machine interaction. This technology translates hand movements into machine instructions, allowing for natural communication. To track motion in these devices, a new approach combines Direction of Arrival (DOA) data with an ultrasonic receiver array that tracks a wearable transmitter. The PDP approach enables precise tracking with a three-sensor receiver array through efficient estimate of 2-D DOA. This method has the potential to enhance airwriting systems' performance.

Alam, Kwon, Md Imtiaz, Hossain, Kang, and Kim [4] provide the TARNet architecture for air-writing recognition, which uses CNN and LSTM as feature generators and recognizers, respectively. The network uses 1-dimensional separable convolution to extract local contextual components from low-level trajectory data, followed by LSTM for high-level dependency detection. TARNet excels for air-writing datasets, achieving accuracy of 99.6% for RTD, 98.74% for RTC, and 95.62% for smart-band datasets. The architecture achieves a compromise between time efficiency and performance, proving effective in various air-writing scenarios. TARNet outperforms individual CNN and LSTM models by providing a comprehensive solution for trajectory-based airwriting recognition.

According to Grigoris Bastas and Kosmas Kritsis [5], air-writing with off-the-shelf smart bands has revolutionized human-computer interaction, eliminating the need for specialized gear. The solution involves recording motion signals with a smart band worn by the user, eliminating the need for extra equipment. The strategy stresses user-friendly interfaces, aligning with current trends. After collecting data from 55 subjects, the user-dependent strategy achieved an average accuracy of 89.2%, while the user-independent method demonstrated robustness with an average accuracy of 83.2%. This study demonstrates how smart bands can build a universally applicable interface for air-writing recognition.The study's findings show that analyzing motion data from smart bands can lead to breakthroughs in gesture-based interaction with computers.

III. Proposed System

In the realm of image processing and pattern recognition, writing in the air has emerged as one of the most intriguing and demanding research topics in recent times. In order to provide software for an intelligent communication technique, the goal of this project is to design a motion-to-text converter. In this case, nonverbal communication in the air is accomplished by using computer vision to identify hand gestures and follow the finger's route. The text that is produced can then be used for a variety of reasons.

Hand gesture tracking and recognition are handled by the MediaPipe framework, and computer vision is handled by the OpenCV package. The application tracks and recognizes hand gestures and tips using machine learning concepts. First, it takes pictures with a webcam and uses a hand tracking module to identify hand landmarks. These landmarks aid in the recognition of hand motions. Then, different activities, like changing colors or cleaning the canvas, can be carried out using the movements that have been identified. In accordance with the gestures recognized, lines are created on the canvas in various colors. Furthermore, a technique for identifying particular gestures—possibly using a machine learning model—allows for more intricate interactions. Overall, the architecture smoothly combines gesture recognition and computer vision algorithms for hand tracking, allowing for real-time interactive sketching on the canvas in response to hand motions made by the user.

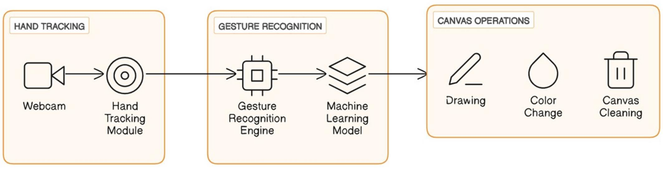

Fig. 1. Block Diagram of system.

A. mediapipe

The mediapipe is employed in the process of hand tracking. For a variety of computer vision tasks, including hand tracking, this library provides easy APIs and pre-trained machine learning models. The algorithm can precisely identify and track human hands in an image or video stream by utilizing mediapipe.

First, some setup is done, such as importing the required files and setting up the webcam's dimensions. The hand tracking module from MediaPipe is imported and set up with the confidence thresholds for tracking and detecting hands.

The webcam continuously reads the Hand Tracking frames. After that, the hand tracking module resizes and runs these frames through its algorithm to identify and follow hands within the frame. On the hands that have been detected, landmarks are located and marked using the findHands method.

After landmarks are identified, more processing is done. For instance, the open and closed positions of fingers are determined by the location of particular landmarks, such the fingertips, which can be helpful for gesture detection and other applications.

B. OpenCV

A variety of computer vision tasks, including image processing, hand gesture detection, and hand tracking, are performed with OpenCV (cv2). There are multiple steps involved in implementation:

Getting the camera (cap) ready to record live video frames is the first step. Every frame is captured and processed using OpenCV. The HandTrackingModule, which is probably a custom module created with OpenCV and Mediapipe, is used to do hand tracking. In order to provide additional hand gesture analysis, this module recognizes and tracks hands in the video feed.

After hands are identified, the code uses landmark positions on the identified hands to identify hand movements, including finger counting. Various finger configurations are associated with distinct actions. In addition, the code uses OpenCV to generate a drawing canvas so that the user can draw with hand motions. The user can draw using a variety of colors, and they can be changed by hovering their hand over the respective buttons.

Using a trained model, character recognition is done in the last section. Preprocessing input pictures using OpenCV includes scaling, grayscale conversion, and thresholding. After the image has been preprocessed, it is fed into a convolutional neural network (CNN) model that has been trained using either the MNIST dataset or a comparable dataset that contains handwritten characters. The input image is used to depict a character that the model predicts the outcome is shown alongside the original image.

C. Numpy

Numerous numerical operations are performed using NumPy (np). The points of various colors for drawing on the canvas are initialized into NumPy arrays. These arrays record the coordinates of dots drawn in each color effectively since they are implemented as deques.As the program runs, fresh points are added to the arrays that represent various colors, and when the deque's maximum length is reached, the old points are eliminated. As a result, drawing points are updated continuously and memory is not used excessively.The given code does not make explicit use of the mathematical functions provided by NumPy. But NumPy offers a wealth of mathematical functions for manipulating and calculating arrays in case more numerical calculations were required.To handle massive amounts of pixel data in photos and movies quickly, NumPy arrays offer effective numerical data storage and retrieval.OpenCV and NumPy arrays work together seamlessly to simplify the translation of image data formats and array representations, which helps with a variety of image processing applications.

D. TensorFlow and Keras

The application uses TensorFlow and Keras to load a pretrained machine learning model. Most likely, this model was trained to identify hand motions or written characters.Once the model has been loaded, the application uses it to forecast the input data. It's probably categorizing hand motions or figures that are created on the canvas in this instance.The input data (maybe character or gesture images) must be properly prepared before predictions can be made. This could entail preprocessing actions like scaling, rearranging, or transforming into the format needed by the model.The prepared input data is subjected to inference using the loaded model. In order to get predictions or classification results, this entails running the input through the model.After then, the model's predictions are analyzed to identify the motions or characters that are known.The user may see this information or it may be processed further.

E. OCR

OCR (Optical Character Recognition) makes use of a CNN architecture that is TensorFlow and Keras implemented. The A-Z Handwritten Data dataset is loaded from a CSV file in order to prepare the data. This dataset is then preprocessed for training.Several Conv2D layers make up this model, which uses convolutional processes to extract pertinent characteristics from the input images. Feature maps are downsampled using MaxPooling layers.Dense and Flatten layers make up the CNN model's last layers. The output of the convolutional layers is transformed into a one-dimensional vector by the flatten layer, which is then used to classify data using densely linked layers. Utilising activation functions like ReLU and softmax, these dense layers generate output probabilities for every class(i.e., the 26 letters of the alphabet).

In the training stage, the model is assembled using a categorical cross-entropy loss function and an Adam optimizer. The model learns to link input photos to their associated character labels across several epochs of training.

F. Dataset

The OCR model was trained using the "A-Z Handwritten Data" dataset, which is included in the supplied code. Handwritten characters from A to Z are seen in these dataset photos.The dataset is commonly kept in CSV (Comma Separated Values) files, where each row represents a picture and the image's pixel values are flattened and saved as columns. A feature (input image) and a label (matching character) are separated out of the dataset.

IV. Results and Discussion

OpenCV, MediaPipe, NumPy and OCR were used in the development of the gesture-controlled whiteboard system, which produced promising outcomes in terms of accuracy, efficiency, and usability.

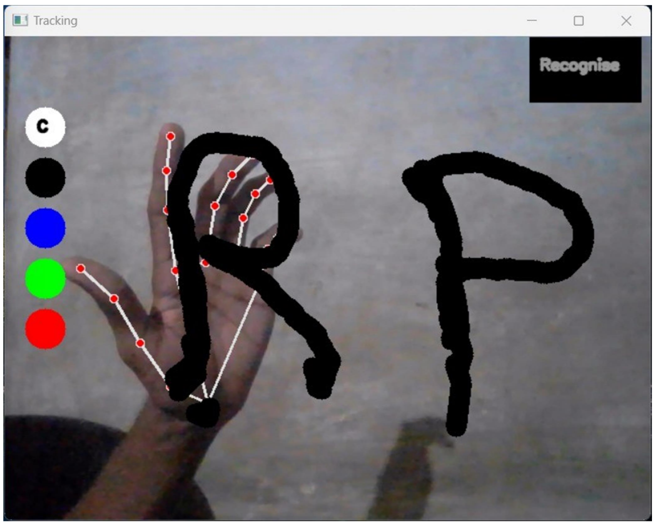

Fig. 2. Capturing video using the webcam.

We tested the accuracy of finger movement tracking and hand detection on a wide range of participants by having them complete a variety of writing and drawing tasks on the virtual whiteboard. The system demonstrated a high degree of accuracy in both hand movement detection and accurate mapping of those movements to the appropriate locations on the whiteboard. Using OpenCV, the hand detection system was able to identify hands in real time with very few false positives or false negatives. Finger positions were precisely monitored by MediaPipe's hand landmark estimation, which made it possible to precisely track finger movements for writing or drawing on the whiteboard.

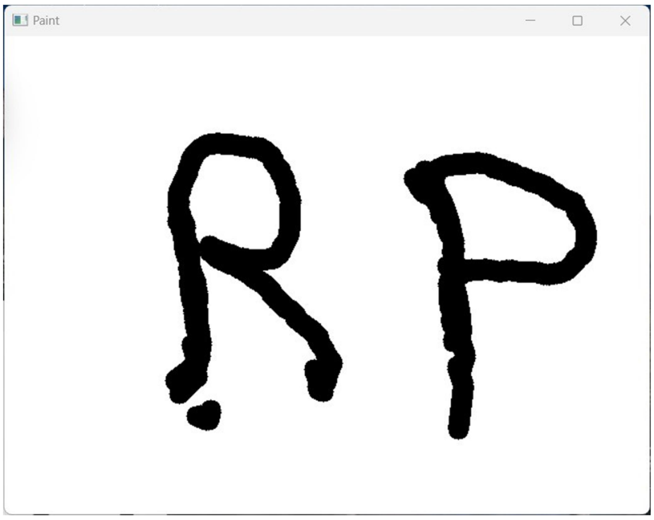

Fig. 3. Performing an action.

A smooth and flawless user experience was guaranteed by the system's real-time responsiveness. By combining OpenCV, MediaPipe, and NumPy, it was possible to process image frames and hand landmark data more effectively and monitor finger movements in real time. There was no discernible lag or delay when writing or drawing on the whiteboard because the system maintained a high frame rate.

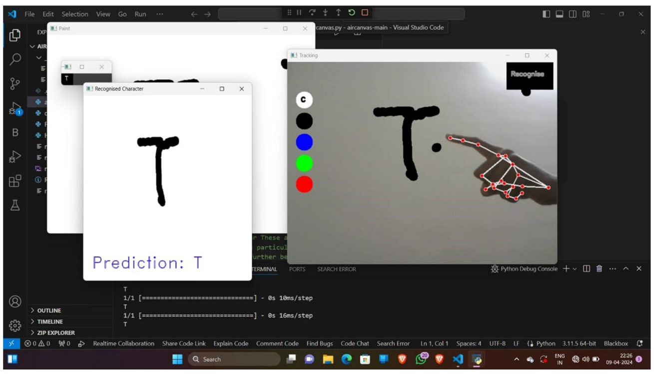

Fig. 4. performing an action.

The system's usability was evaluated in large part through user feedback and usability testing. Participants highlighted how straightforward and simple it is to use, expressing a favorable experience with the gesture-controlled whiteboard. They discovered that the device was a good substitute for conventional physical whiteboards since it was accurate and responsive in capturing their finger movements. The ease with which participants may alter or remove text by merely making an erasing motion with their finger was very appreciated.

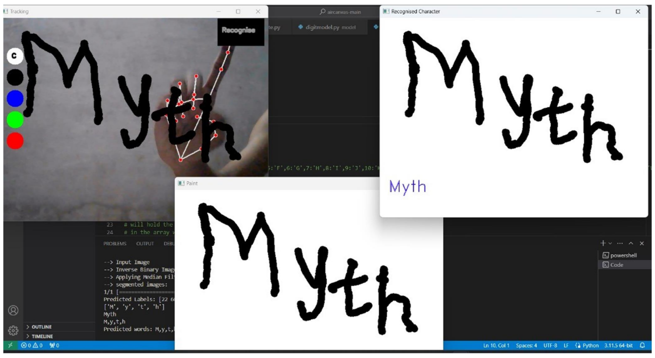

Fig. 5. performing an action.

The system demonstrated a high level of overall user happiness and engagement, suggesting its potential for fostering collaborative and productive communication environments. Utilizing OpenCV, MediaPipe, and NumPy, a gesturecontrolled whiteboard system was created that demonstrates how computer vision and image processing techniques can be used to create engaging and easy-to-use user interfaces. With just a few finger movements, users can write and draw on a virtual whiteboard thanks to the system's hand detection and tracking algorithms. There are many real-world uses for this technology, particularly in collaborative and instructional environments.

The system's overall efficacy is attributed to the precise hand tracking and detection it achieves. The system efficiently recognizes hands in real-time by utilizing OpenCV, guaranteeing a dependable input for tracking finger movements. Accurate mapping of finger positions on the virtual whiteboard is made possible by the precise landmark estimate made possible by the integration of MediaPipe. Furthermore, the OCR (Optical Character Recognition) part of the system exhibits excellent accuracy when transferring handwritten text on the whiteboard to digital format. By enabling users to easily integrate handwritten notes into digital papers, this feature improves usability. Additionally, by improving data processing activities, NumPy's utilization increases the efficiency of the system and guarantees smooth performance—even with enormous datasets. In summary, the gesture-controlled whiteboard system that was created with the aid of OpenCV, MediaPipe, NumPy, and OCR technology shows encouraging outcomes in terms of performance, accuracy, and ease of use. Accurate finger movement tracking, real-time responsiveness, accurate hand detection, and a smooth user experience are all made possible by the effective integration of these opensource frameworks. Furthermore, by precisely transferring handwritten text on the whiteboard to digital format, the OCR component improves usability and makes documentation and cooperation simpler.

With its potential to revolutionize typical whiteboard interactions, this technology could find use in interactive presentations, collaborative workplaces, and educational institutions. Additional enhancements and optimizations may be investigated to augment the system's functionalities and augment its overall efficiency, guaranteeing its efficacy in a variety of settings and circumstances.

A. Perfomance Analysis

The TensorFlow and OpenCV-based real-time hand gesture drawing and recognition program. The HandTrackingModule is used to identify hand gestures, and a pre-trained deep learning model is used to classify them. Based on the gestures that have been recognized, related shapes or patterns are then drawn on a canvas using various colors.

There are various ways to examine this application's performance. First and foremost, hand gesture tracking and detection accuracy and efficiency are critical. It appears that the HandTrackingModule offers a dependable way to identify hand landmarks and follow their motions in real time. However, the accuracy could change based on things like hand orientation and lighting.

Second, the application's general usability is impacted by the gesture recognition accuracy. The user experience is impacted by how well this model recognizes gestures. The dependability of the program may be lowered by problems like incorrect classifications or false positives.

Ultimately, user happiness is influenced by how smoothly and responsively the drawing feature operates. The user experience is improved by the application's capacity to recognize hand motions with accuracy and convert them into flowing lines or forms on the canvas. The performance of the application might be negatively impacted by elements like latency, jitteriness, or drawing delays.

Overall, this application's performance analysis evaluates the responsiveness of the drawing capabilities, the efficiency of real-time processing, and the accuracy of hand motion detection and recognition.

V. CONCLUSION

In conclusion, the Virtual Air Canvas project is a creative work that could fundamentally change how the digital age approaches efficiency, creativity, and communication. It brings a hands-free, immersive experience that surpasses traditional methods and creates a plethora of chances for creative expression and cooperation, redefining online collaboration. The Mark Air project's strength lies in its potential to shape future online collaboration and function as a force for good in a world where remote work and digital communication are now necessities. Its dynamic and seamless virtual whiteboarding environment fosters an environment where ideas are free to flow and enables both individuals and groups to communicate creatively and productively. Furthermore, the project's creative note-taking method streamlines the creative process and demonstrates a commitment to effective and user-centric design, as it allows users to take notes comfortably and handsfree. The Virtual Air Canvas project ensures that collaboration is not only innovative but also inclusive and effective by promoting accessibility as a fundamental value and accounting for varying communication preferences and styles. A testament to the transformative potential of technology, the Virtual Air Canvas project sets a new standard for accessibility in virtual collaboration and makes a significant impact on the digital landscape. Because of its user-centric design philosophy, it is guaranteed to be more than just a tool; rather, it is a platform that fosters productivity and creativity while making collaboration inclusive and simple for everybody.

VI. Future Scope

Although it has a strong basis for a hand gesture detection and sketching application, there are a number of ways to improve its functionality and user experience in the future.

First off, using more sophisticated methods for hand motion recognition could increase the application's adaptability. Although basic gestures like finger counting are recognized by the current implementation, the program would be more effective if the gesture language were expanded to include gestures for certain commands or activities. This can entail using a bigger dataset to build an advanced deep learning model to reliably recognize a wider variety of hand motions.

Secondly, incorporating functionalities that allow for the dynamic modification of drawing settings may grant users greater autonomy over their works. More expressive artwork might be possible, for instance, if users could alter the brush size, opacity, or color mixing mode in real-time with hand motions. It would be necessary to incorporate gesture-based controls for smoothly changing drawing settings into the application's UI in order to provide such functionalities.

Moreover, improving drawing skills through the support of extra drawing tools and features would improve user satisfaction. For example, adding tools like erasing, undo/redo capabilities, and the capacity to create freeform curves or geometric shapes could increase the application's versatility and appeal to users of various creative tastes and skill levels.In conclusion, it is imperative to optimize the application's performance to guarantee seamless and responsive functioning across diverse hardware platforms. This can entail enhancing the code's resource efficiency, taking advantage of hardware acceleration via methods like GPU or parallel processing, and putting in place effective algorithms for real-time gesture recognition and sketching. It would be more accessible and appealing to a wider audience if the application functions flawlessly even on devices with lower specs.

it is imperative to optimize the application's performance to guarantee seamless and responsive functioning across diverse hardware platforms. This can entail enhancing the code's resource efficiency, taking advantage of hardware acceleration via methods like GPU or parallel processing, and putting in place effective algorithms for real-time gesture recognition and sketching. It would be more accessible and appealing to a wider audience if the application functions flawlessly even on devices with lower specs.

The potential to move gesture-based interaction beyond sketching and possibly into virtual reality, games, and interactive presentations. As advances in machine learning and computer vision continue, the program may eventually be able to identify increasingly complex hand movements for a variety of applications. Furthermore, integrating with platforms for mixed reality (MR) or augmented reality (AR) could open up immersive user experiences like virtual designing or sculpting. Multiple users interacting with the canvas at once could be made possible by collaborative features, encouraging innovative teamwork. All things considered, the promise is in utilizing cutting-edge technologies to produce intuitive and captivating experiences outside of conventional input techniques.

Acknowledgment

The authors extend sincere gratitude to Mrs. Sreema E R for her Guidance and to project coordinators, Mrs. Najla Nazar, and Mrs. Gishma K M for mentorship. Their expertise played a vital role in shaping the direction and focus of this research.

References

[1] Taiki .Watanabe, Md. Maniruzzaman , Si-Woong .Jang ,and Jungpil .Shin , "2D camera-based air-writing recognition using hand pose estimation and hybrid deep learning model," 2023.

[2] Saez-Mingorance.B, Mendez-Gomez.J, Mauro.G, Castillo-Morales.E, Pegalajar-Cuellar.M, Morales-Santos.D.P, "Air-Writing Character Recognition with Ultrasonic Transceivers," Sensors 2021.

[3] Chen.H, Ballal.T, Muqaibel.A.H, Zhang.X, and Al-Naffouri.T.Y, "Air writing via receiver array-based ultrasonic source localization," IEEE Trans. Instrum. Meas. 2020.

[4] Tomer Yanay,and Erez Shmueli, "Air-writing recognition using smartbands," Pervasive Mob. Comput. 2020.

[5] Grigoris Bastas, and Kosmas Kritsis, "Air writing recognition using deep convolutional and recurrent neural network architectures,"2020.