Air Canvas

Jagadeesh Vavili, Sai Mohan Reddy Vurabavi, Lokesh Chitturi, Lokesh Palaparthi and Vaibhavee Jani

April 1, 2024

AIR CANVAS

VAVILI JAGADEESH
Computer Science Engineering (CSE),
Parul University, Gujarat, India
vavilijagadeesh@gmail.com

V SAI MOHAN REDDY
Computer Science Engineering (CSE),
Parul University, Gujarat, India
msai51780@gmail.com

CH LOKESH
Computer Science Engineering (CSE),
Parul University, Gujarat, India
lokeshchitturi3@gmail.com

P LOKESH
Computer Science Engineering (CSE),
Parul University, Gujarat, India
lokeshpalaparthi@gmail.com

Prof. VAIBHAVEE JANI, Assistant Professor
Dept. Of Computer Science Engineering
Parul University, Gujarat, India
vaibhavee.jani18427@paruluniversity.ac.in

***Abstract*—In recent years, air based writing has emerged as a compelling area in pattern recognition and image processing, crucial for advancing automation and enhancing man-machine interaction. With a focus on reducing processing time and enhancing recognition accuracy, researchers have explored novel techniques. Object tracking, an integral part of Computer Vision, has gained momentum with faster computers and affordable, high-quality video cameras. This involves three key steps: object detection, frame-to-frame tracking, and behavioral analysis. Object tracking entails addressing issues such as suitable representation, feature selection, detection, and tracking. Its applications span automatic surveillance, video indexing, and navigation. Leveraging this, a project aims to create a motion-to-text converter for intelligent wearable devices, allowing writing in air for various purposes, aiding communication for the deaf and reducing reliance on traditional devices.**

I. INTRODUCTION

Computer vision stands as an interdisciplinary field of study focused on enabling computers to grasp and interpret digital images or videos at a sophisticated level. From an engineering point of view, it aims to understand the tasks akin to those performed by the visual system. Its scope encompasses techniques for analyzing, and interpreting digital images, along with extracting high-dimensional data from the real world to generate numerical or symbolic information, facilitating informed decisions. In practical terms, understanding implies the transformation of visual images (retinal input) into representations of the environment that align with cognitive processes and can prompt appropriate responses Computer vision stands as an interdisciplinary field of study focused on enabling computers to grasp and interpret digital images or videos at a sophisticated level. From an engineering point of view, it aims to understand the tasks akin to those performed by the visual system. Its scope encompasses techniques for analyzing, and interpreting digital images, along with extracting high-dimensional data from the real world to generate numerical or symbolic information, facilitating informed decisions. In practical terms, understanding implies the transformation of visual images into representations of the environment that align with cognitive processes and can prompt appropriate responses.

The scientific discipline of computer vision primarily concerns the principles underpinning artificial systems designed to extract meaningful information from images. Image data can manifest in diverse formats, multiple camera perspectives, data from three dimension scanners, and medical imaging equipment. Computer vision, as a technical domain, strives to implement its concepts and frameworks in the advancement of computer vision systems.

Fig. 1. Computer vision Process

A. *PROBLEM STATEMENT*

The project is aimed at tackling significant societal issues:

1.People with hearing impairment: Many people with hearing impairments communicate using sign language, which most people don't understand without a translator. This creates a barrier to understanding their feelings and emotions.

2.Overuse of Smartphones: Smartphones have become an integral part of our lives, but their overuse can lead to accidents, depression, distractions, and other health problems. Despite their convenience, they can also pose serious risks to our well-being.

3.Paper wastage is a big problem that we often overlook. We use up a lot of paper for writing, drawing, and other tasks. Did you know that it takes about 5 liters of water to make just one sheet of A4 size paper? And a whopping 93 percentage of all paper comes from cutting down trees! Plus, half of the waste generated by businesses is paper, and a quarter of what goes into landfills is paper too.

All this paper waste isn't just bad for the environment because it uses up water and trees, but it also creates tons of garbage that just piles up. But there's a solution called Air Writing that can help. It's a way of writing in the air instead of on paper. This can be especially helpful for people who are deaf or hard of hearing because their air-written messages can be shown using Augmented Reality or turned into speech.

The cool thing about Air Writing is that you can jot things down in the air and then get right back to what you were doing without needing paper. And since everything is stored electronically, there's no need to worry about wasting paper or harming the environment.

II. SYSTEM METHODOLOGY

To get this system up and running, we need a dataset for the Fingertip Detection Model. The main job of this model is to track the movement, specifically the gestures made in the air

A. *Fingertip Detection Model:*

Air writing can be done with a special stylus or air-pens that have distinct colors. However, our system takes a different approach by utilizing fingertip detection. We think people should be able to write in the air without needing to carry around extra tools like a stylus. To make this possible, we've employed Deep Learning algorithms that can identify fingertip positions in each frame, creating a list of coordinates for accurate tracking.

B. *Techniques of Fingertip Recognition Dataset Creation:*

a. Video to Images: Initially, we captured two-second videos of hand movements in various settings. These videos were then split into 30 individual images each. In total, we collected 2000 images. To train our model, we manually labeled this dataset using image labels. Despite achieving an impressive 99 percentage accuracy, the model struggled when presented with backgrounds different from those in the dataset, as the images were all from the same video and environment, resulting in monotony.

b. Distinct Background Photography: To address the limitations of the previous method, we created a new dataset with more diversity. This time, we focused on capturing four distinct hand poses to enable control gestures for the system, as illustrated in Figure 4. The goal was to train the model to accurately recognize fingertips for all four fingers, allowing users to control the system based on the number of fingers shown. For example, one finger for writing, two for converting to e-text, three for adding space, and so on. The dataset comprised 1800 images, and using a script, we auto-labeled it with the previously trained model. After correcting mislabeled images and introducing further refinement, we achieved a 94 percentage accuracy. Importantly, this model performed well across different backgrounds, unlike the previous one.

C. *Fingertip Recognition Model Training*

The main goal of this project is to create and put into action a strong system for verifying the integrity of data stored securely in the cloud. This system will ensure that user data remains intact, confidential, and accessible whenever needed.

III. MOTIVATION

Writing or drawing in the air is a captivating and demanding area of research within artificial intelligence. It's something that sparked our interest, leading us to develop a project called "Air Canvas" – essentially drawing in the air. Drawing or sketching with our hands is something many of us dream of doing. We often imagine doodling freely without any constraints. This project emerged from that very concept. We've created a virtual canvas where you can select colors and draw or write whatever you like simply by moving your hand in the air.

IV. LITERATURE REVIEW

A. *Movement Recognition using LED System:*

This system utilizes a camera to detect the movement of an LED attached to the fingertip. It can accurately recognize patterns drawn on the screen by the LED. The advantage lies in its fast movement tracking and high accuracy. However, it requires the LED to be specifically a red LED, and it may fail if other objects resembling a red LED are present in front of the camera.

B. *System with Air Mouse:*

This device is equipped with sensors that can be worn on the finger, essentially functioning as an Air mouse. It serves as a substitute for a traditional mouse, allowing users to provide various inputs to the system. The key difference is that the Air mouse operates in the air in front of the screen. It detects hand movements such as grabbing or keeping the hand vertical to execute actions like selecting, dragging, or scrolling, similar to a real mouse.

C. *Hand Recognition with Kinect Sensor:*

This system employs a Kinect sensor to capture depth and color information, enabling the detection of hand movements. While Kinect sensors are efficient with large objects, they may encounter challenges in recognizing gestures accurately, particularly with smaller objects. Despite its capabilities, it may struggle with precise gesture recognition for tiny objects.

V. IMPLEMENTATION

A. *Hand Detection*

The process of hand tracking involves the use of computer vision in a computer to recognise a hand from an input image and then maintain attention on the movement of the hand as well as its orientation.

Algorithms

- Import the Libraries
- Initializing the hand's landmarks detection model using Mediapipe
- Read an Image
- Perform Hands Landmarks Detection

output

Fig. 2. Hand Detection Output

B. *Finger Tip Detection*

Finger Tip Detection is the method of using computer vision in a computer to recognise a specific fingertip of the hand from an input image which is captured by the system camera.

Algorithm
1.Import the Libraries

2. Initializing the hand's landmarks detection model using Mediapipe

3. Load an image from the input source using OpenCV or any relevant image processing library.

4. Utilize the initialized hand landmarks detection model to detect the landmarks on the hand within the image.

5. Identify the fingertips by analyzing the landmarks detected on the hand. Determine the positions of the fingertips based on the landmarks provided by the detection mode

output

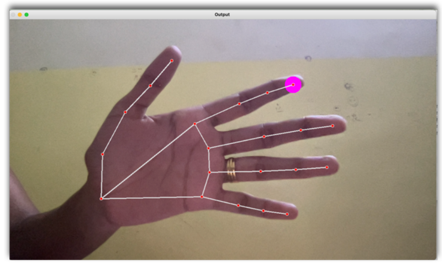

Fig. 3. Finger Tip Detection output

C. *Finger Tip Tracking*

The process of fingertip tracking involves the use of computer vision in a computer to recognise a specified fingertip of a hand from an input image which 23 is captured by the system camera and then maintain attention on the movement of the hand as well as its orientation

D. *Algorithm*

1.Import the Libraries
2. Initializing the hand's landmarks detection model using Mediapipe
3. Read an Image
4. Perform Hands Landmarks Detection
5. Detection of Finger Tips using Hands Landmarks
6. Tracking of detected finger tip

E. *Output*

VI. SOFTWARE REQUIREMENT

A. *python*

Python is a popular programming language known for its simplicity and versatility. It was initially developed by Guido van Rossum in 1991 and has since been supported and expanded by the Python Software Foundation. One of Python's key features is its readability, making it easy for programmers to write and understand code. Its concise syntax allows developers to express complex ideas with fewer lines of code, promoting efficiency and productivity. Python is valued for its ability to facilitate rapid development and seamless integration with various systems.

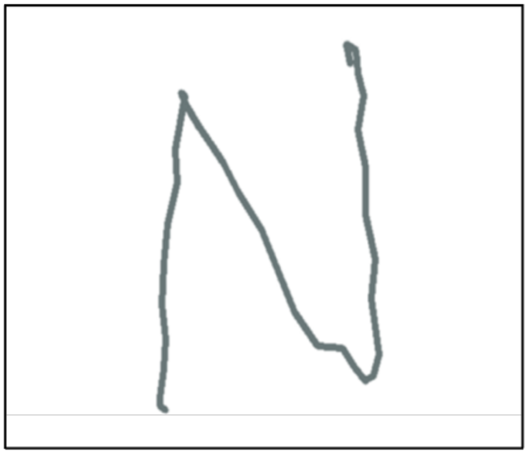

Fig. 4. Finger Tip tracking output

B. *MediaPipe*

MediaPipe, a framework by Google, addresses hand tracking challenges and offers machine learning solutions adaptable to diverse projects. With modules for movement and gesture recognition, it enhances user experience across platforms by detecting hand shapes and motions. This capability forms the basis for hand gesture control, sign language interpretation, and augmented reality overlays. Despite its complexity, MediaPipe Hands provides real-time hand and finger tracking using machine learning, even on mobile devices. This accessibility encourages innovation and exploration in research and development, opening avenues for novel applications and advancements.

C. *OpenCV*

OpenCV is a powerful library used for recognizing and processing images, particularly for tasks like hand tracking and drawing. Its primary focus is on image processing and recognition, making it ideal for various computer vision applications. With methods for object detection and real-time image processing, OpenCV enables the creation of advanced computer vision systems using Python. It's extensively used for tasks like analyzing images and videos, as well as detecting faces and objects with high accuracy.

D. *NumPy*

The ndarray, or N-dimensional array, is a fundamental object within NumPy, a popular Python library for numerical computing. It represents a collection of elements that are all of the same type, arranged in multiple dimensions. Each element within the array can be accessed using a zero-based index. Importantly, all elements in the ndarray occupy the same amount of memory space. Additionally, each element is associated with a data-type object, denoted as dtype, which determines the type of data stored in the array. When accessing elements from an ndarray, Python objects representing the array scalar types are retrieved. This relationship between ndarray, dtype, and array scalar types is illustrated in the accompanying picture.

VII. CONCLUSION AND FUTURE WORK

In the future, we aim to utilize advanced machine learning models to accurately recognize written characters. Our goal is to create a natural interface between humans and systems that eliminates the need for keyboards, pens, or gloves for character input. This involves categorizing letters across various datasets and languages to ensure accuracy. We'll employ sophisticated models to predict alphabets with the highest level of precision. Our approach will revolutionize how characters are identified and inputted into systems, making interactions more intuitive and seamless.

VIII. CONCLUSION

The system we've developed offers a fresh approach to writing, eliminating the need for carrying a phone to jot down notes. It's a convenient way to write on-the-go, especially beneficial for differently-abled individuals, seniors, or those uncomfortable with keyboards. Furthermore, it can extend its utility to controlling IoT devices and even enable drawing in the air. This software could greatly enhance smart wearables, fostering better interaction with the digital world. Augmented Reality adds another dimension, bringing text to life. Despite its strengths, there are areas for improvement. For instance, integrating a handwriting recognizer could speed up writing, while employing hand gestures with pauses for real-time system control could enhance user experience. Additionally, the system sometimes misidentifies fingertips in the background, which could be addressed by refining control gestures. Using a more appropriate dataset and leveraging advanced object detection algorithms like YOLO v3 could further enhance fingertip recognition accuracy and speed. Looking ahead, ongoing advancements in Artificial Intelligence hold promise for optimizing air-writing efficiency even further.//

REFERENCES

[1] Jahnavi S, K Sai Sumanth Reddy, Abhishek R, Abhinandan Heggde, Lakshmi Prashanth Reddy, "Virtual Air Canvas Application using OpenCV and Numpy in Python ", IJCRT, vol. 10, 2022, pp. 855-859.

[2] S.U. Saoji, Nishtha Dua, Akash Kumar Choudhary, Bharat Phogat, "AIR CANVAS APPLICATION USING OPENCV ANDNUMPYINPYTHON", IRJET, vol.08, Aug 2021, pp. 1762-1764.

[3] Joolekha Bibi Joule, Ahsan Raza, Muhammad Abdullah, and Seokhee Jeon, "Tracking of Flexible Brush Tip on Real Canvas: Silhouette-Based and Deep Ensemble Network-Based Approaches", IEEE Access, vol. 8, July 2020, pp. 115782-115784

[4] 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.

[5] S Suraj, Vikas S Shende and P Bharti, "Real Time Object Detection using Deep-Learning and OpenCV

[6] Shahrizat Shaik Mohamed, Nooritawati Md Tahir and Ramli Adnan, "Background modelling and background subtraction performance for object detection"

[7] Mingyu Chen, Ghassan AlRegib and Biing-Hwang Juang, ”Air-writing recognition-Part I: Modeling and recognition of characters words and connecting motions
[8] 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.
[9] 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.
[10] Saira Beg, M. Fahad Khan and Faisal Baig, ”Text Writing in Air,” Journal of Information Display Volume 14, Issue 4, 2013.
[11] Alper Yilmaz, Omar Javed, Mubarak Shah, ”Object Tracking: A Survey”, ACM Computer Survey. Vol. 38, Issue. 4, Article 13, Pp. 1-45, 2006.
[12] Yuan-Hsiang Chang, Chen-Ming Chang, ”Automatic Hand-Pose Trajectory Tracking System Using Video Sequences”, INTECH, pp. 132- 152, Croatia, 2010
[13] 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
[14] 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.
[15] 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
[16] Maryam Khosravi Nahouji, ”2D Finger Motion Tracking, Implementation For Android Based Smartphones”, Master’s Thesis, CHALMERS Applied Information Technology, 2012, pp 1-48
[17] Eshed Ohn-Bar, Mohan Manubhai Trivedi, ”Hand Gesture Recognition In Real-Time For Automotive Interfaces,” IEEE Transactions on Intelligent Transportation Systems, VOL. 15, NO. 6, December 2014, pp 2368-2377
[18] Waung Yang and Shankar Sarkar, ”Coupled grouping and matching for gesture recognition.
[19] Smriti k, Anushka and Vishal Singh, ”Face detection and Recognition: A review”.
[20] Sakeel Ahmad, Noor Najam and Rashid Ahmed, ”Image-based Face Detection and Recognition”