AR Canvas with Python

Piyush Garg[1], Yashika Choudhary[2], Utkarsh Pandita[3], Samarth Singh Thakur[4], Veena Jadhav[5], Dr. Rohini Jadhav[6]

[1,2,3,4] Bharati Vidyapeeth (Deemed to be) University College of Engineering, Pune

[5]Assistant Professor, Bharati Vidyapeeth (Deemed to be) University College of Engineering Pune

[6]Assosciate Professor, Bharati Vidyapeeth (Deemed to be) University College of Engineering Pune

Abstract **Writing is a unified type of communication that allows us to successfully communicate our ideas. Today's standard means of recording information include typing and writing. With a marker or a finger, characters or words are written in the empty area. The pen does not move up and down as it does in typical writing techniques. Human gestures can now control the digital world thanks to the development of clever wearable gadgets. These wearable technologies are capable of recognising and comprehending human activities. Gesture recognition is the process of recognising and interpreting a continuous sequential gesture stream from a collection of input data. Gestures are nonverbal cues that help computers grasp what they're saying. Vision perceives human motions, and computer vision is used to analyse diverse gestures. The project takes advantage of this gap by concentrating on the development of a motion-to-text converter that might be used as software for intelligent wearable gadgets that allow users to write from the air. The technology will employ computer vision to track the route of the finger, allowing for writing from above. The created text may be utilised for a variety of applications, including sending messages and e-mails. For the deaf, it will be a strong way of communication. It's an efficient communication approach that eliminates the need to write, reducing mobile and laptop use.**

Keywords - Character Recognition, Object Detection, Real-Time Gesture Control System, Real-Time Gesture Control System, Smart Wearables.

I.INTRODUCTION

The computer has had a tremendous influence on human existence since its creation. Humans need computer interaction for a variety of purposes, ranging from maintaining records to doing complicated scientific computations. A surge in computer usage has also been attributed to an expansion in the Internet of Things. Around 2000 BC, the art of writing was discovered. Neolithic humans started writing on cave walls. These barriers were quickly replaced by stone, then fabric, and we now report on paper. With recent technology advancements, we are increasingly heading towards a more digital form relying mostly on electronic text materials. We can generate and modify these digital text documents using QWERTY keyboards. Traditional pen and paper writing is gradually being replaced by this electronic keyboard and text.

QWERTY keyboards have also become more portable as a result of touchscreen devices. However, we now consider taking notes on a mobile phone or tablet to be tedious. Furthermore, using one hand to hold a phone prohibits the user from engaging with the actual environment.

With the advent of virtual and augmented reality, the need to develop natural human-machine interaction (HCI) solutions to replace current HCI methodologies is rapidly growing. Automobile interfaces, human activity recognition, and a number of advanced hand gesture recognition algorithms (Molchanov et al., 2015) have all grown in popularity recently; Rautaray and Agrawal, 2015. To type words, however, requires more than just hand motions. In order to achieve a more natural human-computer interface, it will be necessary to develop non-contact air writing devices that can take the place of tactile and electromechanical input fields (HCI). It's not new that utilising a vision-based system to recognise handwriting in the air may be challenging; in fact, a lot of study has been done in this area over the last 20 years. Used a sophisticated device with an infrared and colour sensor in conjunction with their fingers to track and recognise movements of simple geometric objects. A glove and inertial sensors were used to continuously recognise and detect handwriting in the air in (Amma et al., 2012). Misra et al. developed a hand gesture recognition framework that can recognise letters, numbers, arithmetic operators, and 18 printable ASCII characters using a red marker on the tip of the index finger to differentiate fingers. Although older methods using cumbersome motion detection and detecting technologies allowed a range of user behaviour constraints, they nevertheless had high route identification accuracy. For instance, protective gloves could interfere with a user's usual handwriting style and are often seen as an unnecessary burden by many users.

This method eliminates the requirement for mobile phone note-taking. Instead, frequent use of portable tools is necessary to enable airwriting. Finger, movement, and aerial character recognition techniques are used to identify the drawn word. Through the use of OpenCV and CNN, the fingertip is first recognised. The path of this fingertip is then recorded and shown on a dark background. Our aerial photography model is different from our approach for handwriting. This allows for the recognition of handwriting using the model.

II. PROBLEM STATEMENT

There are no highlighters, paints, or relatives in the current technique, which requires simply your fingertips. It's difficult to recognise and characterise an item like a finger from an RGB picture without a depth sensor. Another issue is the pen's lack of up and down action. To write from above, the system employs a single RGB camera. The up and down actions of the pen cannot be monitored since depth detection is not feasible. As a consequence, the complete route of the fingertip is drawn, resulting in an outlandish picture that the model would not recognise. Using real-time hand gestures to modify the system's state requires a great deal of programming attention. In addition, the user must be familiar with a variety of motions in order to effectively manage his strategy. The initiative aims to address certain pressing societal issues.

For starters, persons with hearing loss have several challenges in everyday life. People with this handicap communicate through sign language, despite the fact that hearing and listening are taken for granted. Without an intermediary translation, most nations in the globe will be unable to comprehend your sentiments and emotions. Second, excessive smartphone usage leads to accidents, sadness, diversions, and other disorders that humans are still learning about. Although its portability and convenience of use are highly praised, it has some downsides, including the possibility of death. Paper waste is not commonplace. Scribbling, writing, painting, and other activities waste a lot of paper. The average amount of water used to make A4 paper is roughly 5 litres. Trees provide 93 percent of sources, 50 percent of commercial waste is paper, 25 percent of landfills are paper, and so on. Wasted paper pollutes the environment by wasting water and forests, as well as producing heaps of garbage. These issues may be readily resolved by using air writing. It will be used to communicate with the deaf. Your handwritten writing may be exhibited in augmented reality or converted to voice. One may write swiftly on-air and continue working without being distracted.

Furthermore, there is no need for paper while writing in the air. Everything is digitally preserved.

III. LITERATURE SURVEY

Computer vision and human-machine interaction are two applications of automatic object tracking [1-2]. The literature suggests a variety of uses for tracking algorithms. It was utilised by one group of researchers to interpret sign languages [5, 9], another group for hand gesture recognition [6-7], a third group for text tracking and recognition [8, 18], a fourth group for virtual reality [10] and character recognition based on finger tracking [19- 21], and so on. Bragatto and colleagues devised a system for translating Brazilian Sign Language from video input. For real-time video processing, they employed a multilayer NN (Neural Perceptron) network with a piecewise linear approximation trigger function. The mean complexity time of NN is reduced by using this activation function. They also employed NN in two stages: colour recognition and hand posture assessment procedures. Their findings reveal that the suggested technique works effectively, with a 99.2% detection rate [5]. Cooper also provided a more sophisticated way for handling 3D cell bioprinting than the generic set. Cooper devised a method for reducing tracing by finding flaws in the categorization and tracing procedures in his thesis. Cooper employed two pretreatment methods: one for movement and the other for determining the hand's form. He also utilised the computer to broaden his lexicon. In the pronunciation of a phoneme and the visual representation of phonemes, the viseme is an important posture of the mouth and face. He designs a haphazardly organised learning approach for identifying characters over time. Using Jordan's Recurrent Neural Network, Araga et al. suggested a hand gesture recognition system (JRNN). Their technique used a series of typical static pictures to mimic the 5 and 9 various hand postures. He then proceeds to identify the hand positions using the movie as input. After the posture sequence's temporal behaviour has been detected, JRNN looks for the input gesture. They are also working on a novel training system. For five distinct hand postures, the suggested approach obtains a precision of 99.0 percent, whereas for nine motions, it produces a precision of 94.3 percent [6]. Yang et al., in [7], proposed an alternate method for matching a succession of photos to a pattern, which is common in hand motion recognition. The suggested approach does not rely on skin colour patterns and may be used even if segmentation is inadequate. They used a cross-cluster technique to merge the segmentation and recognition processes. Their findings demonstrate improved performance, with both models losing 5% of their performance. Neumann and colleagues devised a system for detecting and recognising text in real-world photographs. They employed a hypothesis framework that can accommodate numerous lines of text in their essay. They also train the algorithm using synthetic characters, and they employ maximum stable areas (MSER), which are resilient to geometric and lighting circumstances [8].

Wang et al. also explored a color-based motion detection system for both indoor and outdoor settings. To track the item, they employed a camera and a coloured T-shirt in the suggested technique. The suggested method's outcome demonstrates that it may be employed in virtual reality applications. Character recognition systems based on finger tracing have been developed by Jari Hannuksela et al. [19], Toshio Asano et al. [20], and Sharad Vikram et al. [21]. The author of [19] offers a motion-based tracking system that estimates two separate movements by combining two Kalman filtering techniques and expectation maximisation (EM) approaches. Movement of the camera and fingers. The calculation is based on the motion attributes calculated by the scene for each picture. Its fundamental concept is to use a finger to swipe in front of a camera to operate mobile devices. The authors of [20] talk about a visual interface that identifies Japanese katakana letters in the air. They utilised an LED pen and a camera to track the hand's movement. They translate the pencil's signal into direction codes. To reduce the influence of typing speed, codes are normalised to 100 data items, with 46 Japanese characters specified. They obtain a character recognition accuracy of 92.9 percent for a single camera and a gesture directing accuracy of 9 degrees for multiple cameras.

IV. APPROACH

The basic stage in aerial writing is to recognise the location of the writing hand and to recognise it by other motions. Writing in the air is not delineated as a writing sequence, unlike conventional writing, where the pen glides down and up. Events. By counting the number of lifted fingers, the system detects the posture of a writing hand and differentiates it from a non-writing hand.

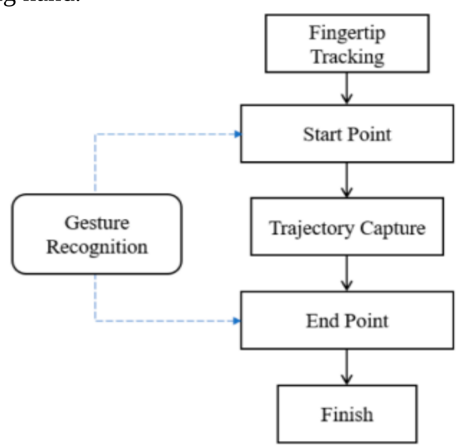

Fig. Flowchart of Proposed System

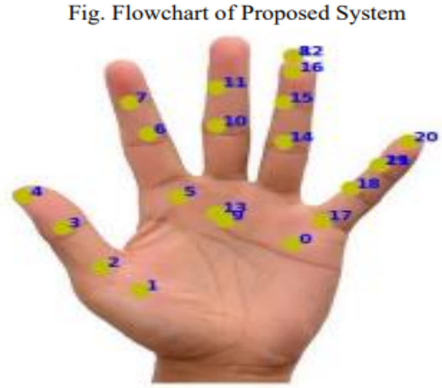

Fig. Hand Region Segmentation

Segmenting Hand Regions

After properly capturing the hand using the aforesaid method, the hand region is segmented in two steps: skin segmentation and background subtraction, with the final binary image of the hand generated as an aggregate of the two. The suggested approach is fast and accurate in real-time segmentation. Although skin hues change widely from breed to breed, it has been shown that skin colour has a narrow range of variation amongst skin types, however skin luminosity varies dramatically.

Background subtraction: Because accurate hand detection with the Faster R-CNN handheld detector is followed by skin colour filtering at the candidate's hand's boundary, the background subtraction step is only used to remove skin-colored objects (not part of the hand) that may be present in the bounding box of the recognised hand.

Hand Centroid Positioning

The system employs two algorithms to obtain the initial estimations of the centre of gravity of the hand, and the final centre of gravity is determined as the average of the two. The initial estimate of the centre of gravity is obtained using the distance transformation technique (xc1, yc1). Each pixel in the distance transform picture is represented by the distance between it and the following edge pixel. The distance between a pixel and its nearest edge pixel was measured using Euclidean distance. As a result, the centre of gravity is determined by the pixel with the greatest intensity in the distance transform picture.

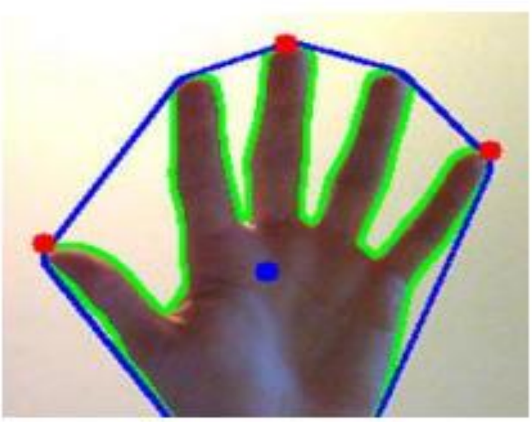

Fig. Hand Centroid Localization

Fingertip Navigation

Fingertip detection and tracking requires hand detection and tracking on successive photos. Experiments reveal that employing the Faster RCNN portable detector for each frame is computationally demanding and results in frame rates that are much slower than real-time performance. To monitor the region of the recorded hand, the KCF tracking algorithm (Henriques et al., 2015) is employed.

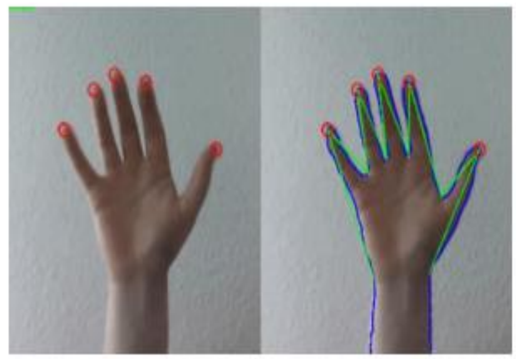

Fig. Fingertip Tracking

Object Color Tracking at the Fingertip

To distinguish hand-colored items, webcam photos must be translated to HSV colour space. The entering picture is converted into HSV space by the algorithm code, which is a highly good colour system for colour tracking.

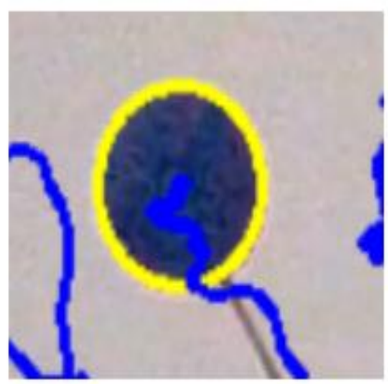

Fig. Colour Tracking of Object at Fingertip

The entering picture is converted into HSV space by the algorithm code, which is a highly good colour system for colour tracking.

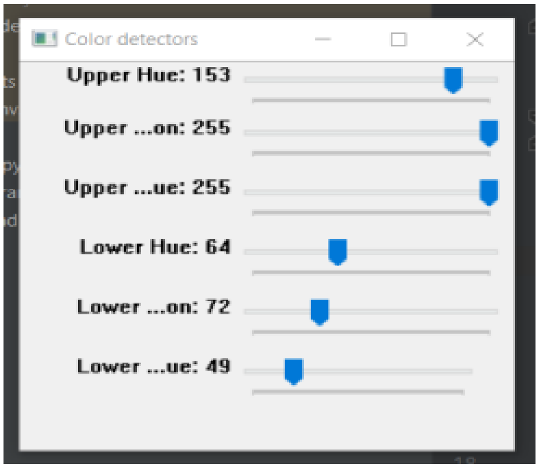

Fig. Colour Trackbar

The HSV values were arranged into the needed colour range of the colourful item you put on your finger by the tracking bars.

Contour Detection of the Color Object Mask

After detecting the Mask in Air Canvas, the next step is to find its centre point in order to create the Line. The system will do certain morphological operations on the Mask in order to clean it up and make it easier to recognise contours.

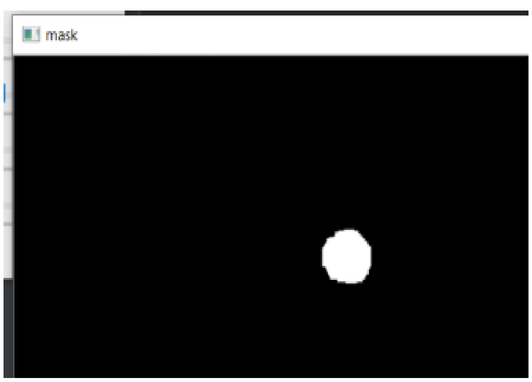

Fig. Mask Detection

Using the contour's location to draw the line

This computer vision project's real logic is to generate a Python deque (a data structure). The deque will remember the outline's location in each following frame, and we'll utilise these collected points to draw a line using OpenCV's drawing capabilities.

Decide whether to press a button or draw on the supplied sheet using the outline position. Some of the buttons are located at the canvas's top. When the pointer reaches this region, the technique in this area is used to activate the pointer.

Fig. Drawing Line using the position of the contour

Image taken from the video stream. Object position: Remove the colour picture from the source: This approach follows the blue-colored index finger movement. We don't have a reference picture, thus each image before it serves as a guide for the next. Take the difference between the photos and extract the object's colour and movement.

Edge Enhancement (EE) - This approach makes the object locating algorithm more resistant to noise, variable lighting conditions, object darkening, and fading, even in low contrast pictures.

V. ALGORITHM OF WORKFLOW

The workflow phase is the most intriguing component of this system. There are several functions involved in writing. As a result, the number of gestures needed to manage this system is the same as the number of activities involved. This system has the following fundamental functionalities:

1. Writing Mode - In this state, the system will trace the fingertip coordinates and stores them.
2. Colour Mode – The user can change the color of the text among the various available colors.
3. Backspace - Say if the user goes wrong, we need a gesture to add a quick backspace

VI. RESULTS

Based on the algorithm defined in the upper section, we will write a code to take input for the system via webcam, and once the color tracking of the object held by the user is done, the designed system can be used.

Fig. System Output

VII. PROJECT SCOPE

This system's primary purpose is to provide a strong means of communication for the deaf, which implies that adopting this project may assist with: An effective communication technique that decreases mobile and laptop use by removing the need to write.

1. An effective communication method that reduces mobile and laptop usage by eliminating the need to write.
2. It helps people with hearing impairments to communicate well.
3. Various purposes, such as sending messages, e-mails, etc., as the generated text can also be used for that.

VIII. CONCLUSION

Traditional writing approaches may be challenged by this technology. It eliminates the need to write down notes on a mobile phone by giving an easy on-the-go solution to do so. It will also be very useful in facilitating communication for those with disabilities. Even older individuals or anyone who have difficulty using keyboards may utilise the system with ease. This technology may soon be used to operate IoT devices, extending its capability. It is also feasible to draw in the air. This solution will be fantastic software for smart wearables, allowing individuals to engage with the digital world more effectively. Text may be brought to life via augmented reality. Airwriting systems should solely follow their master's control gestures and not be led astray by others. Fingertip recognition accuracy and speed may be improved with upcoming object detection techniques like YOLO v3. Artificial intelligence breakthroughs will improve the efficiency of air-writing in the future.

REFERENCE

[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 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
[14] P. Ramasamy, G. Prabhu, and R. Srinivasan, "An economical air writing system is converting finger movements to text using a web camera," 2016
[15] International Conference on Recent Trends in Information Technology (ICRTIT), Chennai, 2016, pp. 1-6.
[16] Surhone, Lambert M., et al. Opencv. 2013. Surhone, Lambert M., et al. Opencv. 2011.
[17] Dawson-Howe, Kenneth. A Practical Introduction to Computer Vision with OpenCV. 2014.
[18] Bradski, Gary, and Adrian Kaehler. Learning OpenCV. 2008,
[19] https://doi.org/10.1604/9780596516130. Howse, Joseph. OpenCV for Secret Agents. 2015. Howse, Joseph. Android Application Programming with Opencv 3. 2015.