Paint / Writing Application through WebCam using MediaPipe and OpenCV

Shaurya Gulati
Department of CSE-AIT, AIML
Chandigarh University
Gharuan, Punjab, India
sgulati1020@gmail.com

Ashish Kumar Rastogi
Assistant Professor
Department of CSE-AIT
Chandigarh University
Gharuan, Punjab, India
ashish.k.rastogi24@gmail.com

Mayank Virmani
Student
Department of CSE-AIT, AIML
Chandigarh University
Gharuan, Punjab, India
virmanimayank15@gmail.com

Rahul Jana
Student
Department of CSE-AIT, AIML
Chandigarh University
Gharuan, Punjab, India
i.rahuljana@gmail.com

Raghav Pradhan
Department of CSE-AIT, AIML
Chandigarh University
Gharuan, Punjab, India
pradhan4334@gmail.com

Chetan Gupta
Department of CSE-AIT, AIML
Chandigarh University
Gharuan, Punjab, India
chetanphenomenal1@gmail.com

2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM) | 978-1-6654-6643-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICIPTM54933.2022.9753939

Abstract: **Writing, a cohesive form of communication that helps us convey our thoughts as efficiently as possible. But in today's modern world, typing is also considered as a way to record information along with writing. Characters or phrases are written withinside the unfastened area with a marker or finger. With the improvement of clever wearable gadgets, the virtual international can now be controlled with the aid of using human gestures. These wearable gadgets can understand and recognize our actions.**

Recognizing and interpreting a non-stop sequential gesture to circulate from the given set of input statistics is gesture recognition. Gestures are non-verbal information used to enhance laptop language understanding. Human gestures are perceived with the aid of using imaginative and prescient, and laptop imaginative and prescient is used to research diverse gestures.

The project takes advantage of this gap and makes a speciality of growing a motion-to-textual content converter that can potentially serve as a program for clever wearable gadgets for writing in the thin air. The device will use the laptop imaginatively and presciently to hint the finger's path, and in that way, it is easy to write from above. The generated textual content also can be used for different purposes, the same as any different content would serve. It will be a whole new means of communication for the deaf as well. It is a powerful communication method that reduces mobile phone and computer utilization with the aid of disposing of what you want to write.

Keywords: WebCam, Media Pipe, OpenCV

I. INTRODUCTION

The computer has had an enormous impact on our so-called simple human lives. These days, humans use computers for most of the reasons, from using it to do their basic work to having done multiple complex tasks. Also, a sudden increase in the IOT (Internet of Things) has led to a rise in computer use, especially in this ongoing pandemic situation where everything is currently being managed with the help of the Internet and computers.

The art of writing has changed drastically from medieval times when at first, humans used to write on animal skins and then replaced it with cloth and at last, we came up with paper. But in the ongoing pandemic situation, we are not supposed to write on paper and pass that paper to people who are not even next to us. We became more acquainted with typing and digital documents and thus replacing our traditional pen and paper.

Also, as we are getting more and more advancements in technology as we are going, also having advancements in Virtual and Augmented reality, the way we interact with the technology, as well as computers, is getting changed as well with the change in HCI. Thus with the oncoming of new technologies, the need to replace the traditional human-computer interaction with unconventional methods is also increasing rapidly.

In particular, the interfaces related to the interplay reachable gestures have gained their recent popularity in many application areas, such as Automotive interfaces, human interest recognition, and numerous advanced procedures to hand gesture recognition have been advanced. However, only hand gestures aren't enough for getting content into text. This tells us for the improvement of non-touch in-air writing application systems that can replace tactile and electromechanical input fields that are used in touch screens, leading to an extra herbal human-computer interplay (HCI) approach. There has been a lot of work done in the past years in the field of AR/VR and thus, capturing handwriting in the frames of a video is not an extensive task now. One of the primary works with the aid of using a sophisticated device with an infrared and color sensor to track and discover trajectories made by the fingertips using this device. Also, in some case studies inertial sensors with a glove were used to discover and apprehend handwriting from the air continuously. Recently, an advanced hand gesture recognition framework was capable of spotting alphabets, numbers, mathematical operators, and 18 ASCII characters and it used a purple marker put on the tip of the index

978-1-6654-6643-1/22/$31.00 ©2022 IEEE

finger to pick out motion made by the finger. Despite quality overall performance in phrases of course recognition accuracy, previous procedures the use of bulky motion detection and detection hardware renounced several obstacles in person behavior. For example, carrying protective gloves can disrupt the person's herbal handwriting sample and is often viewed with the aid of using many users as an undesirable burden.

Finger and movement recognition techniques and aerial character recognition are used to recognize the drawn word. We use OpenCV and CNN in our model. The fingertip is identified first. Then, the trajectory of the fingertip is recorded and drawn on the background that is provided. It can later be modeled into a handwriting recognition model as well if required.

II. PROBLEM STATEMENT

The existing system only works with fingers and there are no highlighters, paintings or relatives. The main challenge was to identify and characterize the finger from a RGB image and work with it without using the depth sensors. Another problem is the lack of up and down movement of the pen. The system uses a single RGB camera to write from above. Since depth detection is not possible, the top-to-bottom activities of the pen are unable to be tracked. Therefore, the entire way with the fingertip is drawn, and the resulting image comes out to be unrecognized and absurd by the model. Also using real-time hand gestures to move the system from one state to another or from one option to another requires a lot of code care. In addition, the user needs to have information about the gestures in order to control his plan well. The project focuses on solving some critical social issues.

First, people with hearing problems have many problems in daily life. Although listening and listening are taken for granted, people with this disability communicate using sign language. Most countries in the world cannot understand your feelings and emotions without an intermediate translator. Second, the excessive use of smartphones causes distractions in teenagers to cause anxiety and psychological problems in adults and other illnesses that humans can still discover. Although its portability and ease of use are deeply admired, it has its fair share of drawbacks as well. It will serve as a communication tool for the deaf. Your text written in the air can be presented in AR or converted to speech. One can write on-air quickly and keep working without any distractions.

III. LITERATURE REVIEW

Automatic object monitoring has many uses in computing, such as Computer vision and human-machine interaction [1-2]. Various applications of monitoring algorithms are suggested in the literature. Many groups of researchers used it for different purposes some translated signal languages [5, 9], some used it for text and hand gesture popularity [6-7], and some used it for text monitoring [8, 18], and to analyze the pixel movement of objects for virtual reality [10] and individual popularity primarily based totally on finger monitoring [11-13], Bragatto developed a technique that took video as input and automatically converted to Brazilian Sign language. They used a multilayer NN (Neural Perceptron) community with a piecewise linear approximate trigger function for real-time video processing. This activation function reduces the suggested complexity time of the neural network. In addition, they used NN in two stages: the first step being color popularity and the second step being the assessment of the hand posture. A detection charge of 99.2% was found for the suggested technique [5]. Cooper also introduced a technique for dealing with 3-D mobile ular bioprinting; this is more complicated than the generalized set. Cooper developed a way that reduces tracing with the aid of figuring out mistakes in his thesis's classification and tracing processes. Cooper used two pretreatment steps; one is for movement, and the other is used to decide the form of the hand. He extensively utilized the display screen to enlarge his vocabulary. The viseme is a vital function of the mouth and face in the pronunciation of a phoneme and the visible illustration of phonemes. Over time, he develops a poorly dependent getting to know technique that identifies characters.

Araga proposed a hand gesture reputation device using Jordan's Recurrent Neural Network (JRNN). Their device modeled distinctive hand positions through a series of consultant static images. He then takes the pictures as enter and starts to classify the hand poses. JRNN unearths the entered gesture after the positional and behavior of the posture series has been recognized. The proposed technique shows a precision of 99.0% for 5 distinctive hand positions, while it achieves an accuracy of 94.3% for 9 gestures [6]. In [7], Yang et al.; mentioned an opportunity technique to match a series of images to a pattern, which commonly happens with hand gesture reputation. The proposed technique isn't primarily based totally on pores and skin color styles and can paint in spite of bad segmentation. They blended each segmentation technique and popularized the use of a cross-cluster technique. Their results show higher overall performance with an overall performance lack of 5% for each model. Neumann et al. evolved a technique to discover and recognize text in authentic images. In their article, they used a speculation framework that can deal with a couple of traces of text. They additionally use artificial characters to educate the algorithm, and, finally, they use maximum stable areas (MSER), to provide the robustness to the factors of geometric and lightning fixture conditions. [8].

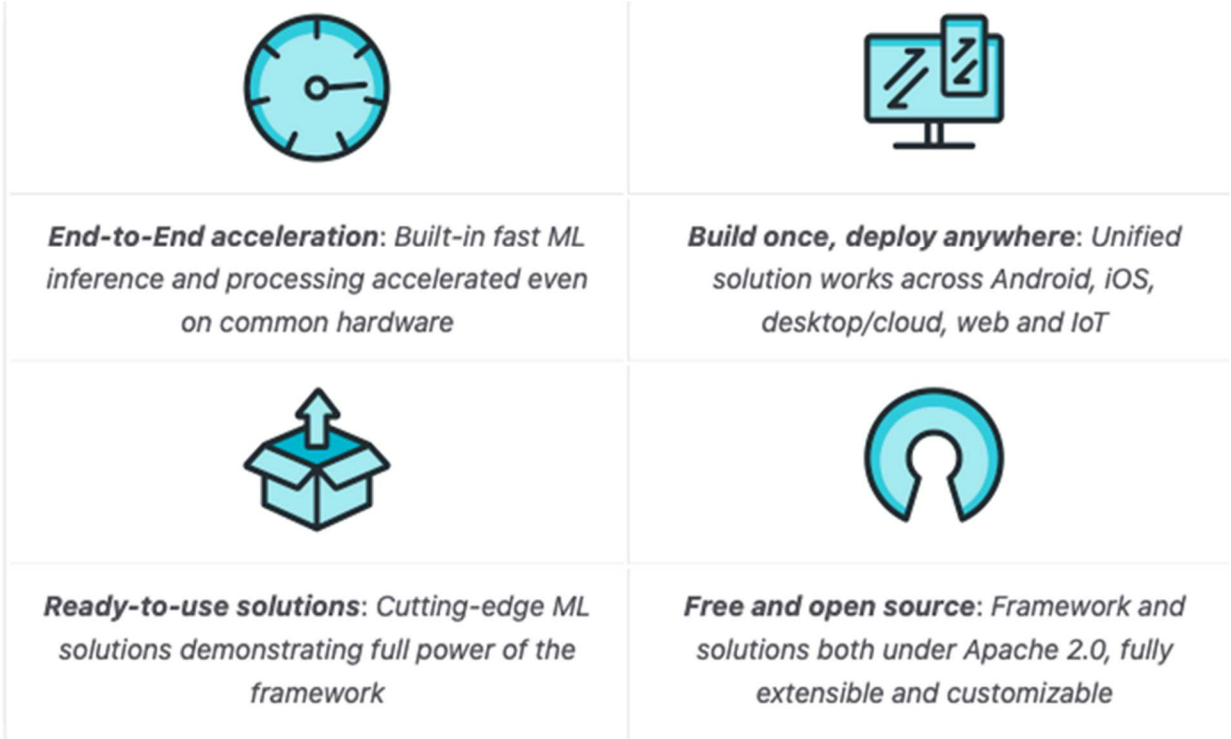
Furthermore, Wang mentioned the color-primarily based movement detection system to distinguished environments such as indoors and outdoors. In the proposed approach, they used a webcam and a coloured T-blouse to tune the object of focus. The result of the proposed approach shows that the

proposed approach can be used for virtual applications. Jari Hannuksela [11], Toshio Asano [12], and Sharad Vikram [13] have character recognition systems primarily based totally on finger tracing. In [11], the writer gives a movement-primarily based total tracking algorithm that combines the two techniques, Kalman filtering strategies and expectation maximization (EM) methods to estimate two different motions; Finger and digital digicam motion. The estimate is primarily based totally at the movement residences that the scene calculates for each image. Its main idea is to control cell devices by truly swiping a finger in front of a digital digicam. In [12], the authors talk about a visible interface that acknowledges Japanese katakana characters in the air. To observe the motion of the hand, they used a light/ LED pen and a digital digicam. They converted the sign of the pen into directional codes. Codes were normalized to a hundred information gadgets to get rid of the impact of typing speed, in which forty six Japanese characters are defined. The character recognition accuracy of 92.9% was achieved with a single digicam and with multiple cameras, the accomplished gesture directing accuracy was of 9°.

IV. Proposed System And Methodology

For hand detection, rather than building specific models with compromising functionality and accuracy, we used MediaPipe so that we do not have to compromise in any aspect, be it be accuracy or functionality.

MediaPipe: MediaPipe offers machine learning solutions to many problems that are cross-platform and customizable as well.



Since we needed to record hands and their movements we used MediaPipe Hands, i.e., MediaPipe Library specifically built for the Hands and its tracking.

MediaPipe Hands: To improve the user experience and its ease of use, MediaPipe hands library was used since it has one of the best abilities to perceive the shape and motion of the hands.

MediaPipe Hand is a Machine Learning primarily working in computer vision field, it can be used for hand tracking, pose detection, face masking etc. Although the other approaches are available as well but they rely primarily over the power of the systems and powerful environments, this method can even work over a mobile device because of its ease of use and portability. It can even be used over multiple hands and its creative use cases stimulating new uses in new researches are emerging as quick as ever.

By using MediaPipe, we will track the movement of our index finger. If we are using index and middle finger then we are in selection mode, we can select any color by simply hovering over the specific color. After selection is done we will simply use just the index finger to draw. The draw mechanism works by drawing individual points at a very small interval of time that those points appear to be a line.

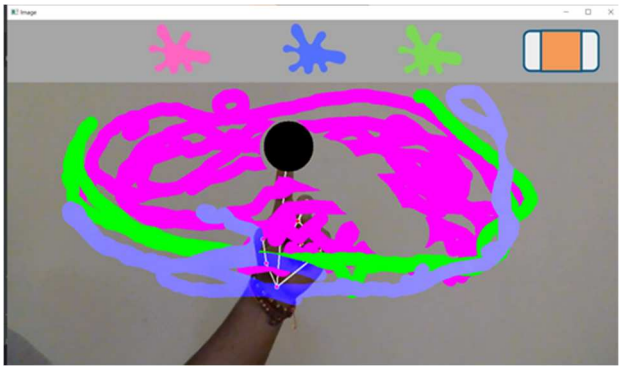
While selecting and drawing modes we will store the location of our hand pointer with the color we are using. Using numpy we will make a plain white canvas of (1280, 720) pixels on which we will draw, the output that we get on white canvas we will superimpose it on the original frame video window.

V. RESULT

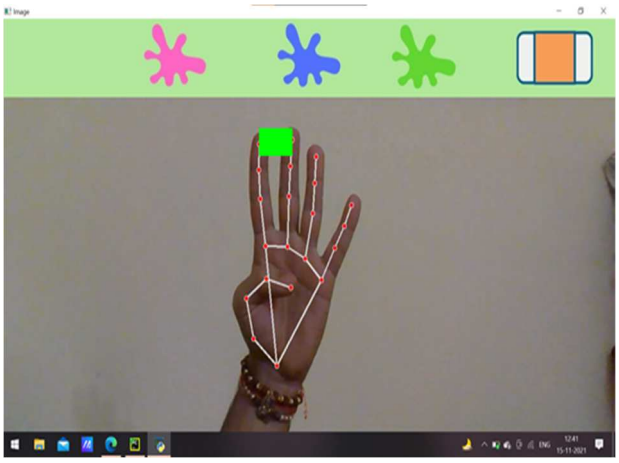
Based on the algorithm defined in the upper section, we wrote a code to take input for the system via webcam, and once the color tracking of the object held by the user is done, the designed system can be used.

The user can choose between 3 colors and an eraser. After choosing the color, the user can draw or write with the respective color.

The live video that the webcam takes or records in makes up the background of the canvas. It can also add up to the use of writing over some picture or diagrams.



Also, We created a gesture for clearing the whole screen at once by using all 4 fingers so that we do not have to erase everything one by one.



VI. CONCLUSION

This system has the potential to challenge traditional writing methods. Eliminates the need to carry a cell phone in your hand to take notes, providing an easy way to do the same on the go. It will also be very useful in helping particularly skilled people to communicate easily. Even the elderly or those who have difficulty using keyboards can use the system effortlessly. By extending the functionality, this system can also be used to control IoT devices in a short time. Drawing in the air can also be made possible. This system will be great software for smart wearable devices through which people can better interact with the digital world. Augmented reality can bring the text to life.

Aerial writing systems should only obey the controlling gestures of their master and should not let the people around you fool you. Upcoming object detection algorithms like YOLO v3 can improve the accuracy and speed of fingertip recognition.

VII. FUTURE SCOPE

The scope of this system is mainly used as a powerful means of communication for the deaf, which means implementing this project can help.
An effective communication method that reduces mobile and laptop usage by eliminating the need to write.
It helps people with hearing impairments to communicate well.
Teaching can be done in Virtual Classes through this without the need for writing pads.
AR and VR technologies can get a huge leap of scope in the upcoming future.

REFERENCES

[1] Y. Huang, X. Liu, X. Zhang, and L. Jin, "A Pointing Gesture Based Egocentric Interaction System: Dataset, Approach, and Application," 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Las Vegas, NV, pp. 370-377, 2016.

[2] P. Ramasamy, G. Prabhu, and R. Srinivasan, "An economical air writing system is converting finger movements to text using a web camera," 2016 International Conference on Recent Trends in Information Technology (ICRTIT), Chennai, pp. 1-6, 2016.

[3] Saira Beg, M. Fahad Khan and Faisal Baig, "Text Writing in Air," Journal of Information Display Volume 14, Issue 4, 2013

[4] Alper Yilmaz, Omar Javed, Mubarak Shah, "Object Tracking: A Survey", ACM Computer Survey. Vol. 38, Issue. 4, Article 13, Pp. 1-45, 2006

[5] Yuan-Hsiang Chang, Chen-Ming Chang, "Automatic Hand-Pose Trajectory Tracking System Using Video Sequences", INTECH, pp. 132- 152, Croatia, 2010

[6] Erik B. Sudderth, Michael I. Mandel, William T. Freeman, Alan S. Willsky, "Visual Hand Tracking Using Nonparametric Belief Propagation", MIT Laboratory For Information & Decision Systems Technical Report P-2603, Presented at IEEE CVPR Workshop On Generative Model-Based Vision, Pp. 1-9, 2004

[7] T. Grossman, R. Balakrishnan, G. Kurtenbach, G. Fitzmaurice, A. Khan, and B. Buxton, "Creating Principal 3D Curves with Digital Tape Drawing," Proc. Conf. Human Factors Computing Systems (CHI' 02), pp. 121-128, 2002.

[8] T. A. C. Bragatto, G. I. S. Ruas, M. V. Lamar, "Real-time Video-Based Finger Spelling Recognition System Using Low Computational Complexity Artificial Neural Networks", IEEE ITS, pp. 393-397, 2006

[9] Yusuke Araga, Makoto Shirabayashi, Keishi Kaida, Hiroomi Hikawa, "Real Time Gesture Recognition System Using Posture

Classifier and Jordan Recurrent Neural Network", IEEE World Congress on Computational Intelligence, Brisbane, Australia, 2012

[10] Ruiduo Yang, Sudeep Sarkar, "Coupled grouping and matching for sign and gesture recognition", Computer Vision and Image Understanding, Elsevier, 2008

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

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

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

[14] A.D. Gregory, S.A. Ehmann, and M.C. Lin, "inTouch: Interactive Multiresolution Modeling and 3D Painting with a Haptic Interface," Proc. IEEE Virtual Reality (VR' 02), pp.4.