Air Handwriting using AI and ML

Prof.Shital Patil1, Homesh Chaudhari2, Homeshwari Chaudhari3, Komal Kapase4, Saurabh Shinde5

*1 Assistant Professor, Department of Information Technology, Sir Visvesvaraya Institute of Technology, Nashik, Maharashtra, India
*2,3,4,5 Department of Information Technology, Sir Visvesvaraya Institute of Technology, Nashik, Maharashtra, India

---***---

Abstract - Air-writing refers to virtually writing linguistic characters through hand gestures in three dimensional space with six degrees of freedom. In this paper a generic video camera dependent convolutional neural network (CNN) based air-writing framework has been proposed. Gestures are performed using a marker of fixed color in front of a generic video camera followed by color based segmentation to identify the marker and track the trajectory of marker tip. A pre-trained CNN is then used to classify the gesture. The recognition accuracy is further improved using transfer learning with the newly acquired data. The performance of the system varies greatly on the illumi nation condition due to color based segmentation. In a less fluctuating illumination condition the system is able to recognize isolated unistroke numerals of multiple languages. The proposed framework achieved 97.7recognition rate in person inde pendent evaluation over English, Bengali and Devanagari numerals, respectively. Object tracking is considered as an important task within the field of Computer Vision. The invention of faster computers, availability of inexpensive and good quality video cameras and demands of automated video analysis has given popularity to object tracking techniques. Generally, video analysis procedure has three major steps: firstly, detecting of the object, secondly tracking its movement from frame to frame and lastly analysing the behaviour of that object. For object tracking, four different issues are taken into account; selection of suitable object representation, feature selection for tracking, object detection and object tracking. In real world, Object tracking algorithms are the primarily part of different applications such as: automatic surveillance, video indexing and vehicle navigation etc. The generated text can also be used for various purposes, such as sending messages, emails, etc. It will be a powerful means of communication for the deaf. It is an effective communication method that reduces mobile and laptop usage by eliminating the need to write.

Key Words**:** Air Writing, Character Recognition, Object Detection, Real-Time Gesture Control System, Computer Vision , Hand tracking.

1.INTRODUCTION

Air-writing systems render a form of gestural humancomputer interaction. Such sys tems are especially useful for building advanced user-interfaces that do not require tradition mechanisms of linguistic input such as pen-up-pendown motion, hardware input devices or virtual keyboards by providing an interface for writing through hand gestures in three dimensional space with six degrees of freedom. The input scheme in such systems is fundamentally different from a generic pen-up-pen-down input mechanism due to the fact that in case of the former there is no robust way of defin ing start (pen-down) and stop (pen-up) states while writing. Unlike conventional writing air-writing systems lack actual anchoring and reference position on writing plane, gestures are guided by considering imaginary axes in three dimensional space. Essentially these facts contribute to the increased variability of writing patterns in such scheme and thereby account for the non-trivial nature of the problem. With rapid development of depth sensors in recent years such as Kinect and LEAP Mo tion, the possibility of air-writing systems has been emerged. Depth sensors along with computer vision techniques are used to track finger tips followed by recognition of the performed gestures using a trained model. But these sensors are not widely available to common devices which restricts these systems to be easily accessible. While depth sensors are not widely available, generic camera is embedded in many commonly used devices. Therefore a generic video camera based air-writing system could be greatly beneficial. But unlike depth sensors a generic camera does not pro vide information regarding scene depth or bone joints and therefore making it more difficult for processing and achieving reliable recognition accuracy.

2. Literature Survey

A computer system can interpret and comprehend hand movements or gestures using a technology called hand gesture recognition. It facilitates communication between machines and people without the use of tangible objects like keyboards or mice. Hand gesture recognition systems typically use computer vision techniques and machine learning algorithms to interpret and classify hand gestures. Real-time hand gesture recognition involves various techniques that allow for human-computer interaction (HCI). Researchers have developed different approaches to recognize hand gestures in real-time, such as using marker identification and tracking technology, data glove-based methodology, skin color identification algorithm. Real-time hand gesture recognition involves various techniques that allow for human-computer interaction (HCI). Researchers have developed different approaches to recognize hand gestures in real-time, such as using marker identification and tracking technology, data glove-based methodology, skin color identification algorithm, LED light source tracking, ultrasonic distance sensors, and skeleton-based models [1]. Hand gesture recognition has numerous applications. Some of them are listed below: A. Human-computer interaction It enables intuitive and natural interaction with computers, smart TVs, gaming consoles, and virtual reality systems. B. Sign language interpretation Hand gestures can be interpreted to facilitate communication with individuals who use sign language. C. Robotics and automation Hand gestures can be used to control robots or machines in industrial or domestic settings. D. Augmented reality Hand gestures can be used to manipulate virtual objects or interact with augmented reality environments. In one approach [2], researchers have used two colored markers worn on the tips of fingers to produce eight hand gestures, which are then sent to a desktop or laptop computer equipped with a camera. They have employed the "Template matching" technique to find markers and the Kalman Filter to track them. In another method, a data glove with ten soft sensors measures joint angles of five fingers to record gesture information. Techniques like gesture spotting, gesture sequence simplification, and gesture recognition are used to recognize gestures in real-time. A skin color identification algorithm [3] is used in another approach to translate American Sign Language (ASL) from real-time video into text. This method employs two neural networks, SCD and HGR, to extract features from the image pixels. Two different methods, Pixel segmentation and finding the fingertip, are used to extract the characteristics. Another method [4] suggested another solution in this context that allows users to create the alphabet or type anything they desire by just sweeping their finger over a vibrant LED light source. To extract the movement of the finger, only the color of the LED is tracked. For 3D hand gesture detection, researchers have used a skeleton-based model that obtains an effective descriptor from the Intel Real Sense depth camera's hand skeleton linked joints. Roy et al. [5] introduces a system that combines augmented reality (AR) and traditional airbrush techniques to enhance computer-aided painting experiences. The authors present the Augmented Airbrush, a physical airbrush tool that incorporates digital overlays and virtual painting effects. The system leverages augmented reality technology to superimpose virtual elements onto the artist's view, allowing them to see a combination of the real-world environment and digital objects. In conclusion, these different approaches to hand gesture recognition enable various real-time applications, such as virtual paint applications and airbrush models that enhance the HCI experience [6]. This paper proposed a solution for Air paint application using Computer Vision techniques.

3. METHODOLOGY

Air-writing is one of the most popular forms of writing. Writing words or letters with finger or hand movements in free space is called air-writing. The recognition of air-writing is a particular case of gesture recognition in which gestures can be matched to write digits and characters in the air Move your hands in a way that co-ordinates with the words you are speaking. Act naturally. Waving your hands and arms erratically is not natural behavior—most people don't do it in everyday conversation, so don't do it during a presentation. Consider the gestures you make when you talk to family and friends. Convolutional Neural Network: Convolutional Neural Networks specialized for ap- plications in image and video recognition. CNN is mainly used in image anal ysis tasks like Image recognition, Object detection Segmentation. There are Four types of layers in Convolutional Neural Networks: 1) ConvolutionalLayer: In a typical neural network each input neuron is connected to the next hiddenlayer. In CNN, only a small region of the input layer neurons connect to the neuronhidden layer. 2) Pooling Layer: The pooling layer is used to reduce the dimensionality of the fea-ture map. There will be multiple activation pooling layers inside the hidden layer ofthe CNN. 3) Flatten:-Flattening is converting the data into a 1-dimensional array for inputtin gitto the next layer. We flatten the output ofthe convolutionallayers to create a sin glelong feature vector. 4) Fully-Connected layer: Fully Connected Layers form the last few layers in the network. The input to the fully connected layer is the output from the final Pooling or Convolutional Layer, which is flattened and then fed into the fully connected layer.

The proposed machine right here we will construct an air canvas which could draw some thing on it by means of simply shooting the movement of a coloured marker with a digital camera. Right here a colored item on the tip of the finger is used due to the fact the marker. In this specific paper we Could be using the computer vision strategies of Open CV using python [5] .Here color detection And tracking are used with a view to obtain the objective of the paper. Color detectors are used to detect the colors used in the air canvas. In this firstly the colored bead is inentified, and Bead is tracked . marker is defined by its color then track it by defining the boundaries of the color to track our marker. Through this we can paint and write in open space [6,7]. Here we shall input live feed and capture the movement of our marker. So, now by just waiving your finger in the air you can draw on the screen using different colors. Coordinates are being detected using the bead so that if it goes in the range of blue color or red or yellow or green it changes to the color and the text is represented in the respective color chosen. The process includes reading the frames and converting captured frames to HSV color space .The input that is the text or images you have drawn in the air are then converted into commands and perform related task in real time. The advantage consists mostly in the ability to track the particular marker that corresponds to each pointer on the air canvas [8]. It keeps track of the marker's coordinates and then displays the colours on the screen as you write [9,10]. It also has the ability to draw in a variety of colours that are present on the air canvas, and these colours may be adjusted to the user's preferences in a comfortable manner. This also has the benefit of being able to clear the screen immediately by just selecting the clear all box when it appears on the display, which is another bonus. Additionally, it motivates the user to enter text immediately after the canvas is filled with text and to desire to write anything new even though the canvas is already full of text.After the paper programme is finished, all of these activities may be performed by simply sketching in the air, and there is no need to make any kind of physical touch with the computer at any point. Therefore, it shouldn't be too difficult for any user to gain access to this.

User interface: User have to have the color bead on their finger as the bead color which has been initialized in the program. Now the user needs to make sure that there is no other color as the bead color in the background which may cause some disturbance. Now making sure above guidelines to be followed the user can access the air canvas with much ease.

Web camera accessing: It is crucial to have access to a live feed from the camera while working with the air canvas because the paper itself is the primary medium for live classes and instruction. As a result, the user has the option of using the camera that is already installed on the laptop, an external web cam, or anything else that is suitable to their needs. Because of this, the web camera needs to be turned on in order to access or analyse the movements of the user's finger while they are using this. Drawing is done by waving, It is crucial to have access to a live feed from the camera while working with the air canvas because the paper itself is the primary medium for live classes and instruction. As a result, the user has the option of using the camera that is already installed on the laptop, an external web cam, or anything else that is suitable to their needs. Because of this, the web camera needs to be turned on in order to access or analyse the movements of the user's finger while they are using this. Drawing is done by waving, which ensures that there is as little direct physical contact with the computer as is humanly feasible; this is why we name it "air canvas." Because of this, the user is able to write on the canvas by monitoring the bead or the colour marker that is now attached to their finger.

Detecting coordinates: After the WE camera has been enabled, the waving can be detected using the bead that has been placed on the finger; this allows the bead to be followed, which ultimately results in the formation of the coordinates in the form of a mask. When the finger is waved around in the appropriate manner, the coordinates for each of them are identified so that the text can be rendered in the appropriate format according to the preferences of the user.

Color detection: Because writing or sketching with multiple colours will be simpler, this paper screen primarily contains red, blue, green, and yellow. These colours were chosen because they are easy to understand while using different colours. Therefore, the names of the colours are displayed at the top of the screen alongside a representation of the colours so that the user may easily use the colours in the manner that best suits their needs. Therefore, each colour has its own unique set of coordinates, and when a colour marker enters the range of a particular colour, that color's coordinates are applied, and the marker's colour is changed to match the colour of the selected colour.

Screen wiping:It's possible that the screen, which has been filled in or scribbled on, needs to be wiped clean. In order to perform such an operation, the option to clear all rectangle boxes has been provided here. This is located at the very top of the screen and has been given the coordinates that are identical to those used for the colours. The user's input will cause the typeface to be removed, and the screen will be cleared if it is within the range of the clear all option.

Displaying the screen: Through the use of the live stream, the user is able to view the canvas. By accessing a web camera, detecting a colour maker, and providing a bead for the formation of the mask, then writing by waving fingers via which coordinated are being detected, and the colour being identified by the given coordinates, everything is presented on the screen.

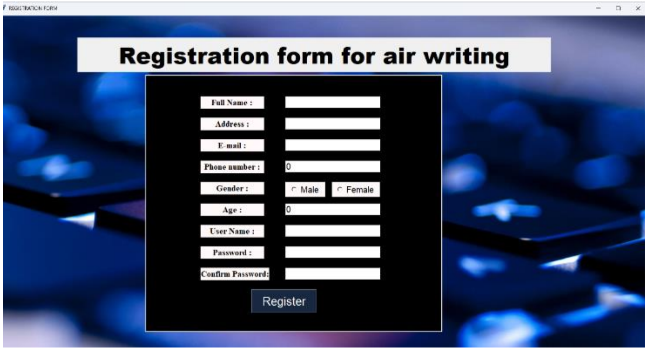
4. PROCESS OF PROJECT

The proposed work utilized machine learning-based approach called the "BlazePalm" and "HandLandmark" models to perform hand detection and landmark localization, respectively. These models are trained on large datasets containing labelled hand images. It consists of two modules namely hand detection module and localization module. These two modules are explained in the below sections. A. Hand Detection (The model is responsible for detecting the presence and rough position of hands in an image. It uses a deep neural network architecture called a Single Shot MultiBox Detector (SSD). The SSD architecture combines a base network (usually a convolutional neural network) with additional layers that predict bounding boxes and class probabilities. Each bounding box that is detected is classified by the model as either containing or not containing a hand. The BlazePalm model uses the network to interpret an input image during inference. The result is a collection of bounding boxes, several of which contain hands. These bounding boxes offer a first estimation of the positions of the hands. B. Hand Landmark Localization (HandLandmark) The precise landmarks or keypoints on the identified hands are localised by the HandLandmark model. It employs a Convolutional Pose Machine (CPM), a convolutional neural network variant. To forecast the coordinates of particular landmarks, like finger tips, joints, and the centre of the palm, the CPM model is trained. During inference, the HandLandmark model takes the cropped hand region (as determined by the bounding box) and processes it through the network. The output of the model consists of the predicted coordinates for each landmark. The detected landmarks are then post-processed to map them back to the original image coordinates. The Mediapipe library provides utility functions to handle this mapping and drawing the landmarks on the image. The mathematics involved in training these models includes techniques like backpropagation, gradient descent, and loss functions to optimize the model's parameters. The training data is annotated with ground truth bounding boxes and hand landmark coordinates, which are used to compute the loss during training. The paper implementation involves a virtual canvas which contains 5 tabs at the top of it such as, line, rectangle, draw, circle, and eraser. At first the canvas will detect our hand and then tracks it when we select 1 tab out of 5 then it will perform the function as per the tabs. Like, if we want to draw a circle then, we can point our index finger at circle tab through which our tab gets selected and then, with index and middle finger both we can draw the circle.

5.IMPLEMENTATION

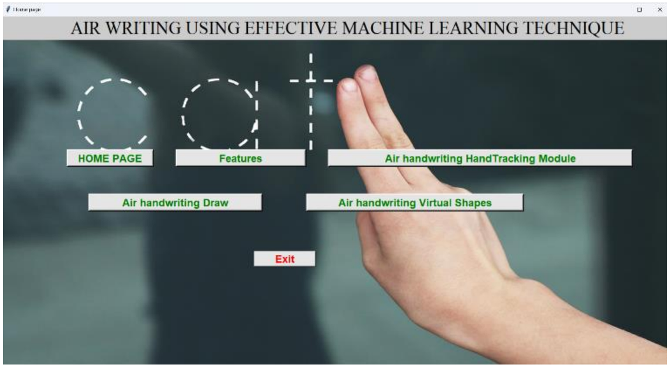
1. First Screen:

2. Registration:

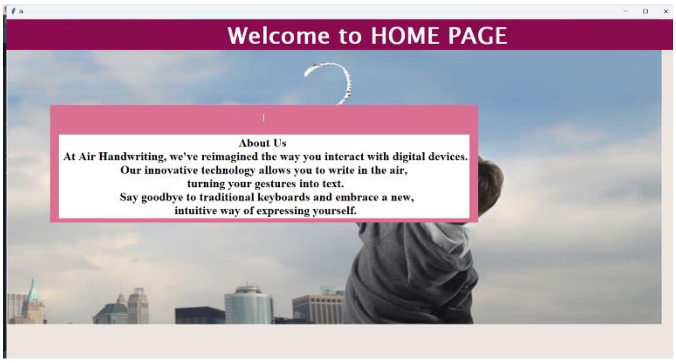


3.Login:

4.Get Started:

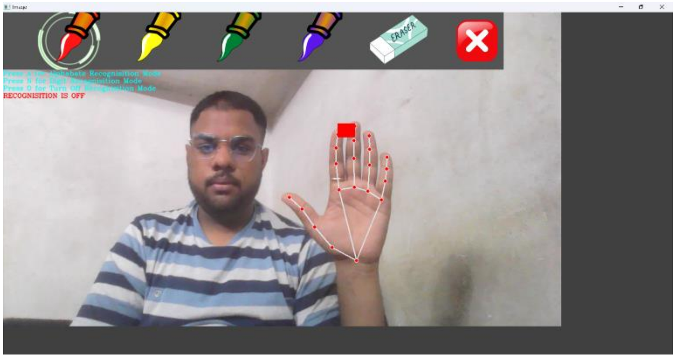


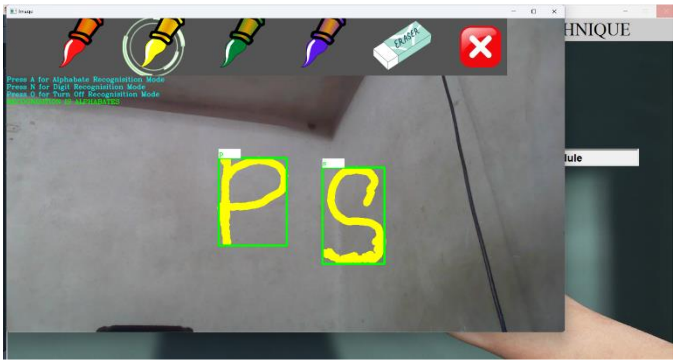
5.Home page:



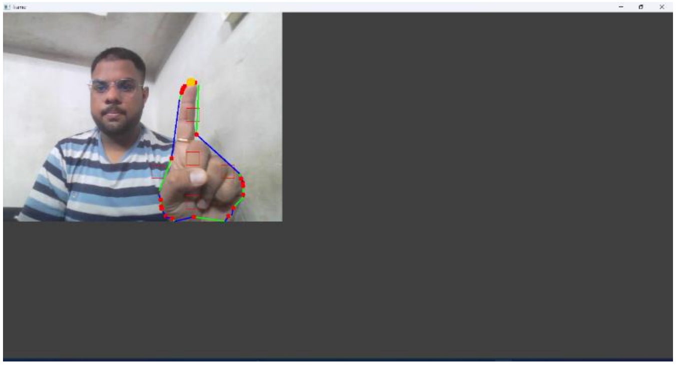
6.Features:

7:Air HandWritting Tracking Module:

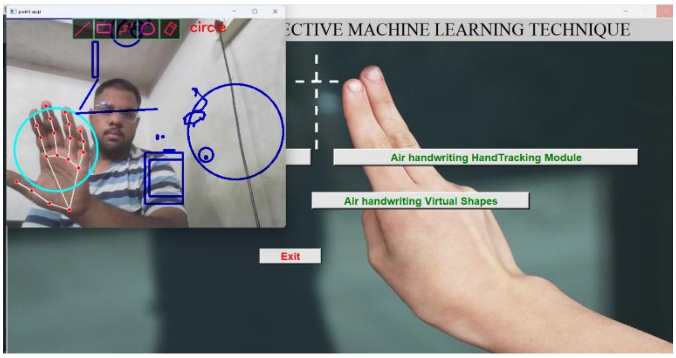




8:Air Handwritting Draw:



9.Air Handwritting virtual shapes:.



CONCLUSIONS

The system has the potential to challenge traditional writing methods. It eradicates the need to carry a mobile phone in hand to jot down notes, providing a simple on the go way to do the same. It will also serve a great purpose in helping especially abled people communicate easily. Even senior citizens or people who find it difficult to use keyboards will able to use system effortlessly. Extending the functionality, system can also be used to control IoT devices shortly. Drawing in the air can also be made possible. The system will be an excellent software for smart wearables using which people could better interact with the digital world. Augmented Reality can make text come alive.

ACKNOWLEDGEMENT

The implementation and documentation of this project would not be succeeded without the kind supportfrom individuals. First of all, we would like to express our special gratitude to **Prof. Shital.S.Patil** who always gives us valuable advice and kind assistance to completethis project. Last but not least, we would like to thank the Faculty of Information Technology, Savitribai Phule Pune University for giving us the great knowledge. Finally, we would like to give our appreciation to our parents who support us since the beginning till the endof this project.

REFERENCES

1. P. Molchanov, S. Gupta, K. Kim, and J. Kautz, "Hand gesture recognition with 3D convolutional neural networks," in Proceedings of the IEEE CVPR workshops, 2015, pp. 1–7.

2. P. Molchanov et al., "Short-range fmcw monopulse radar for hand-gesture sens ing," in IEEE RadarCon. IEEE, 2015, pp. 1491–1496.

3.J. Lien et al., "Soli: Ubiquitous gesture sensing with millimeter wave radar," ACM Transactions on Graphics (TOG), vol. 35, pp. 1–19, 2016.

4. X. Lou et al., "Gesture-radar: Enabling natural human-computer interactions with radar-based adaptive and robust arm gesture recognition," in SMC. IEEE, 2018, pp. 4291–4297.

5. X. Cai et al., "Efficient convolutional neural network for fmcw radar based hand gesture recognition," in Adjunct Proceedings of the ACM UbiComp/ISWC'19, 2019, pp. 17–20.

6. S. Hazra and A. Santra, "Robust gesture recognition using millimetricwave radar system," IEEE sensors letters, vol. 2, no. 4, pp. 1–4, 2018.

7. S. Hazra and A. Santra, "Short-range radar-based gesture recognition system using 3D CNN with triplet loss," IEEE Access, vol. 7, pp. 125 623–125 633, 2019.

8. Y. Sun et al., "Real-time radar-based gesture detection and recognition built in an edge-computing platform," IEEE Sensors Journal, 2020.

9. A. Santra and S. Hazra, Deep Learning Applications of Short Range Radars. Artech House, 2020.

10. M. Arsalan and A. Santra, “Character recognition in air-writing based on net work of radars for human-machine interface,” IEEE Sensors Journal, vol. 19, no. 19, pp. 8855–8864, 2019.