



VIRTUAL PAINTING USING HAND GESTURE DETECTION AND VOICE COMMAND RECOGNITION DEVELOPED USING PYTHON

¹Kavya J , ²Lavanya K , ³Prakruthi P , ⁴Prarthana MP , ⁵Dr. Saravanan Kumar R

^{1,2,3,4} Student, Department of CSE, Dayananda Sagar Academy of Technology and Management, Bangalore, India.

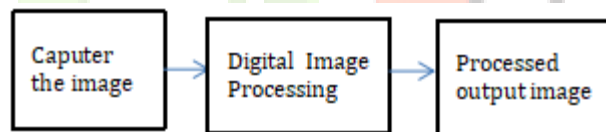
⁵ Professor, Department of CSE, Dayananda Sagar Academy of Technology and Management, Bangalore, India.

Abstract: There exists many tools that help painting in real life but painting or drawing by making hand gestures in air and by voice commands is relatively a new and exciting approach. It provides the user with a dust free environment to draw and illustrate his thoughts to the recipient. The virtual paint is developed using libraries of python like OpenCV, MediaPipe and speech recognition module which act as a backbone of Augmented reality. This program uses hand gestures and speech command from the user to interpret his thoughts on desktop screen.

Index Terms - Python, Mediapipe , OpenCV, Hand Gesture Recognition, Voice Commands , Augmented Reality

I. INTRODUCTION

Digital Image Processing deals with manipulation of digital images through a digital computer. DIP focuses on developing a computer system that is able to processing on an image. The input of that system is a digital image and the system process that image using efficient algorithms and gives image as an output. In today's world digital image processing is almost used in every field and is going with time to time and with new technologies. For example, the digital image processing is used in the field of Medical, Robot vision, Pattern Recognition, Video Processing and image sharpening and restoration.



The language required for this project is python, which has a more comprehensive library and easy to use syntax and understanding of the basics and can be implemented in any open-source language cv support. Virtual painter is developed with the help of OpenCV MediaPipe and speech recognizing ability of the Python. OpenCV and MediaPipe are the libraries of the Python are mainly used for Image Processing. These libraries are used to track the movement of an object in front of Webcam. Using this tracking feature, the user can draw on the screen by moving his index finger for drawing and two fingers for selection in the air, in front of the webcam. This is interesting and challenging as this involves real-time webcam data generated by tracking the movement of the finger of the user to draw and make selection. The marker is chosen as the monochromatic object on your fingertip. In order to achieve the goal, the color tracking and detection process was used for this project. The colored markers used here are detected and a mask is generated. In our project we make use of morphological processes which processes images based on shapes. The process involves modifying an input image by structuring it and then generating output image. Two basic morphological processes are: Erosion-It is used to eliminate away the boundaries of foreground object and is used to diminish the features of an image. This is achieved by removing pixels from the boundaries of the image. Dilation-It is mostly used to increase the object area and to make the features get elevated. This is achieved by adding more pixels to the boundaries of the image. The Virtual Painter is mainly used to interpret the user thoughts and also helps to improve the digital painting skills of the user by providing the more efficient and sophisticated tool that enables the user to easily undo the actions which cannot be done in physical paintings. It is used in many fields such as Education-helps the professor to explain the concepts to the students in a more easily understandable manner, Medical-helps the doctor in conveying the types and effects of the diseases the patient has, Business-Helps the presenter to easily convey the main implementation of the project to the client, Entertainment-It can also be used by any person with a laptop for entertainment purpose to enhance his painting skills.

II.LITERATURE SURVEY

[1] Naive Bayes Approached in Color Detection using Pandas & OpenCV developed by Hassan Jillani Farooqi, Anum Anwaar Chauhan and Azfar Wajid Siddiqui used in detecting color using Naive Bayes algorithm. Color detection is used to identify the name of the color. The task of identifying the color is easy for humans whereas it is difficult for a computer to identify the color. They have employed

Naïve Bayes algorithm, Pandas library and OpenCV libraries that are implemented in python language. The simplest technique that we have for building classifiers could be Naïve Bayes classifier. It works on a given problem instances by designating class labels to instances which is an array of feature values. The classification labels will be taken from a finite set. The basic assumption of Naïve Bayes Classifier is that each feature is independent of each other features, given the classification labels. The OpenCV (OpenSource Computer Vision) library is mainly devised for efficiency of the computer with a special function in real-time applications.

[2] SCLAiR: Supervised Contrastive Learning for User and Device Independent Airwriting Recognition developed by Ayush Tripathi, Arnab Kumar Mondal, Lalan Kumar and Prathosh AP mainly focuses on writing in air with the movement of finger which is the most recent technology in today's world. The concept of Air writing recognition could be thought of as an issue of detecting letters in the empty area with the movement of finger. It could be understood as a more specific case of gesture recognition in Processed output image Digital Image Processing which the library of gestures maps to letters which is present in any specific language. Since most of the people nowadays prefer to carry all light weight objects in their hands such as smart wearables air writing recognition which uses the motion sensors in the smart watches will be the means through which the input is collected from the user for the applications that involve humancomputer interaction. This tool mainly aims at developing air writing recognition which is user independent and is mainly based on supervised contrastive learning. The classification method involves two phases. In the first phase, given an encoder network which needs to be trained with supervised contrastive loss. In the next phase, the encoder weights which are preserved helps in training the classification head. This tool has been tested with both supervised and unsupervised settings.

[3] Multiple Object Detection using OpenCV on an Embedded Platform developed by Souhail Guennouni mainly focused on object detection. A high response to cover the terms of image processing, object detection of the embedded system. Object detection technology has much interest on wider applications. Using the techniques of advancement, the video surveillance and the processing of image plays a major role. They had used OpenCV libraries. The complex aspects were solved using cascade classifier using large scale systems. In order to achieve reliable and accurate near real-time object detection applications, they were keen on high performance of the complex and crucial fields. Related to security concerns in different fields, the application can be deployed in different platforms. It's easier to use mobile devices equipped with cameras and processors. Therefore, the proposed application in real time gives better results of more complex instances in a simpler way.

[4] Face Detection and Recognition Using OpenCV by Maliha Khan, Sudeshna Chakraborty, Rani Astya, Shaveta Khepra is based on OpenCV. Various face algorithms used under OpenCV are Haar cascade_frontend.xml, Eigenfaces see create EigenFace Recognizer (), Fisher(Placeholder3)faces see createFisherFacerecognizer(), Local Binary Patterns Histogram. Haar Cascade is defined as a series of square shaped functions which together form a group of wavelets or a base. This is then used to detect objects. Under the Eigen faces approach we must normalize the project on Eigen Space. This is used for recognizing face and position on face detection applications. OpenCV is organized into five main elements. The MILL portion includes statistical Classifiers and Clustering Tools. The HighGUI portion includes the Graphical User Interface, Image and Video input/output. The CXCORE includes the basic structures and algorithms, XML support, drawing functions. One of the main constraints of Face recognition is that the picture must be planned before using the tool. Instead of the resulting image projection into face space, we can create the network from the data as an output. For detection algorithm Karhunen-Loeve is made use.

[5] Machine learning in fine art paintings developed by Y. Lu and Y. Wang mainly focused on capturing of images for paintings. Captured images for painting was difficult to explore as paintings had contained complex expressions and datasets with zero annotations. Therefore, the painting captioning became very hard to perform task. Photographic images were better than picture captioning. This is an attempt to generate contents of descriptions in an innovative way. The two solutions for the above problems were mainly focused on small datasets of image captioning with detailed text and diagrams. We trained the model to generate the ways to image captioning without photographic images to improve meaningful flow. Compared to CNN and LSTM models, baseline model for solving the erroneous descriptions. This research takes a fresh approach to the challenge of fine art painting for image captioning. They created a model using virtual networks and train the model with complete fulfilling annotated descriptions.

[6] A software that recognizes sign language was developed by Nimisha K P and Agnes Jacob. It is developed using different software like ANACONDA, MATLAB. The two main approaches to construct this software involves (i) image based and (ii) sensor based. The main approach of image based is it involves one or more cameras to capture an image sequence of the signer performing the sign and then uses processing to recognize the sign. In the sensor based approach an instrumental glove with embedded sensors is used to track the hand gestures. The input is in the form of series OS captured images and the output will be in the form of speech, which helps the normal people to understand the sign language of the deaf and dumb. This project involves the following process image acquisition, image processing, feature extraction, sign language classification and finally sign language translation. For feature extraction the following methods are taken into considerations Histogram based features, Principal Component Analysis (PCA), Convolution Neural Network (CNN), Region of Interest Segmentation (ROI). For classification the following algorithms are taken into considerations Artificial Neural Networks (ANN) which is a convolution network used for classification of objects. It learns through set of training data model. And the other classification tool is Support Vector Machine (SVM), which is an ML tool for supervised learning. It classifies the data according to the probability if the data.

[7] A smart VIRTUAL MOUSE USING HAND GESTURE developed by Abhilash S S, Lisho Thomas, Naveen Wilson, Chaithanya C is an efficient way of making use of hand gestures. The hand gesture is the most effortless and natural way of communication among people when they cannot explain something in terms of words. This project presents a new way of controlling cursor system by means of computer vision based This tool makes the computer cursor move according to the user hand gesture. There exist different signs for left click and dragging of the cursor. The hand bearing color caps or toaes are made use for this project. In this project the user's hand bearing color cap is tracked in two dimensions. It is implemented using python and its one of the important libraries called OpenCV which is the back bone of AR. The OpenCV is a python library that provides important function for digital image processing which makes the work of the developer easy. The input is taken in the form of sequence of image frames captured via webcam. The web camera captures the frames in every instance during the entire course of the program. The captured image is analyzed and then processed

using various algorithms. The image which is in RGB color space is converted to HSV color space. There are more than 150 color-space conversion methods available in OpenCV. But only two methods are considered they are BGR to Gray and BGR to HSV. Then masking technique is used to process some of the region of the image. After these processes a bitwise AND operation is performed on the input image with the Threshold image which results in red colored object to be highlighted. The resultant image is stored in res folder. The output is displayed on the monitor.

[8] Combining Hand Detection and Gesture Recognition Algorithms for Minimizing Computational Cost which is developed by Roman Golovanov, Dmitry Vorotnev and Darina Kalina mainly focuses on combining both hand detection and gesture recognition that reduces the cost of computation. The crucial part of Human-Computer interaction (HCI) is hand recognition. The more usual method to develop a recognition system is by using a pre-trained CNN (Convolution Neural Network). There have been a new design structures called convolution pose machine stand in for skeletal model of hand or human body that is captured in the form of an image with an accuracy of high level. These design structures need very high cost and efficient computational resources which are not accessible most of the time. It uses the combined system of hand gesture recognition where it makes use of a hand detector to identify the hand that is presented on a frame and then it shifts to gesture classifier if the hand is been identified. This system is experimented on gesture bases that are accessible in public and also on video sequences. It majorly aims at developing the system that involves the recognition of hand which makes use of less required computer resources in order to minimize the cost required to implement this system.

[9] An efficient tool for Online Teaching using OpenCV developed by Gangadhara Rao Kommu makes use of morphological operations such as erosion and dilation. In the first step the frames are read and then converted to HSV color space. This step is performed because it is easier for color detection. In the next step the canvas is prepared. The ink buttons for different color options and other selection options are displayed on the canvas. Further the mask is preprocessed with morphological operations. In the subsequent step they detect the contour. The contour's center coordinates of the largest contour is found. This information must be stored in the array for next frame analysis. Lastly the points that are stored in the array are drawn on the frame and the canvas. This system provides two types of display options for the user. One is the white board display and the other being the camera visibility display. The advantage of this system is that the user can change the colors and also clear the screen. The enhancement that can be added to this project is the option of choosing different font size options for drawing. The main motivation for this project is to enhance the classroom teaching experience. This system also promotes dust free classrooms. The system enhances the learning experience as students can visualize the lessons in an innovative manner. The major advantage of this tool is to make the work of a teacher easy and effective which will reduce their work without spending much funds. It was very much useful during the pervasive outbreak of the world-wide spread of the contagious disease.

[10] Virtual Painting with OpenCV using Python by Yash Patil, Mihir Paun, Deep Paun, Karunesh Singh, Vishal Kisan Borate have made use of OpenCV modules for developing a virtual painter. The modules they have mainly made use of are: CORE module that has the data structures and fundamental functions performed by other modules. IMGPROC module contains image processing related information. VIDEO module contains detection of motion and object tracking algorithm. ML module contains machine-learning interface. HighGUI module contains basic I/O interfaces and multi-platform windowing capabilities. The process starts by importing the necessary libraries. BGR values are assigned to the ink colors. Further a window is created that provides various options. The video of the user's movement is captured using a webcam. The video is split into frames and then stored. Different Morphological Operations are performed on the video frames. Two basic morphological operations include erosion and dilation. Finally, the output is displayed on the screen. The ultimate goal of this project is to develop an application that promotes Human Computer Interaction (HCI). This system can be classified into two steps Extraction method and Feature Estimation. The extraction method starts with segmenting the object. Next the object must be tracked. Images can be easily segmented based on color. The enhancement for this project includes that the system must be compatible even with mobile phones. The application must consume less power as mobile is powered only by a battery.

III. RESEARCH METHODOLOGY

The input is taken in the form of frames of images from the video captured by the webcam. The image is then processed and the required part of the image is extracted by using various image processing algorithms. The Extracted part is then matched with the trained model. The desired output is then displayed on the screen. In case of Voice Commands, the speech is recorded through Microphone of the device and then the required feature is extracted. Finally, the comparison is made with the trained model and the result is displayed on the screen. Voice Commands are only used for selecting tools from the toolbox. The following flowchart explains the various steps involved in developing of Virtual Painter.

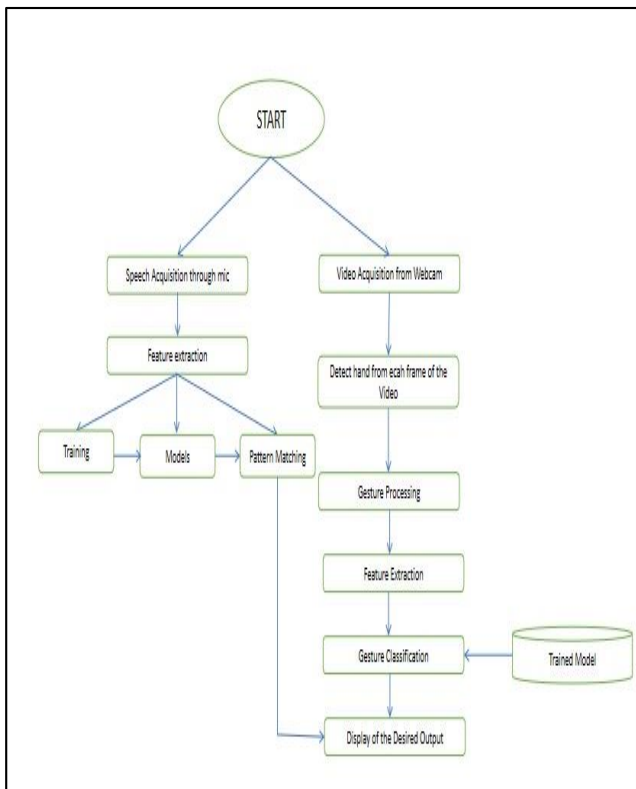


Fig 1: Flowchart

1. Acquisition: The image frame is acquired from the video captured by the camera or other sources. The webcam or the laptop camera is the main source for this project. The captured image is unprocessed at this stage and it is sent to the system. The speech command is collected from the mic connected to the device and it is also unprocessed. It is sent to the system for further processing.

2. Preprocessing: Image processing is a method that involves certain operations on the image in order to get an enhanced image or to extract some useful information from it. It is performed on the unprocessed captured image. The main advantage of preprocessing is that it reduces unwanted noises from the image. It is achieved through smoothing and blurring. The speech preprocessing involves filtering the speech signals from the ambient and undesirable noises, detecting speech activity and normalizing the length of the vocal tract.

3. Feature Extraction: Feature Extraction is a method that is used to reduce a large input dataset into required relevant features. This is achieved by reducing large input data into smaller meaningful groups for processing. Feature Extraction in terms of image refers to the process of transforming raw data into numerical features that are helpful while processing. Convolution Neural Network (CNN) is used for extracting required features from the image. In case of speech feature extraction is accomplished by altering the speech waveforms to the form of parametric representation at a relatively lesser data rate for further processing and analysis.

4. Classification: Image Classification is the fundamental task that tries to categorize the recognized sign with the particular label. The particular sign is recognized by comparing the sign with the trained data set. There are several image classification algorithms that is used to categorize the sign. Speech Classification involves categorizing the filtered part of the speech according to the trained model by means of comparison.

5. Displaying of Output: The output is displayed on the monitor of the device. The output is the drawing made by tracking the user hand gestures. This involves the above steps as described. The user can draw using his index finger and can make selection using middle and index finger at the same time. The voice commands are used only for selection of tools. The user is provided with options of changing brush size, eraser size, color and also can pick the predefined shapes provided in the tool section.

IV. CONCLUSION

The Virtual Painter is made capable of detecting the hand gestures and voice commands and was trained to draw according to the user instruction via hand gestures and speech commands. This project involves various libraries of the Python such as OpenCV and MediaPipe for image processing and speech recognition module of the Python for voice command recognition. This tool provides the user to work in a hassle-free environment such as dust and painting stains. This allows the user to interpret his thoughts in a contactless environment.

REFERENCES:

- [1] Naïve Bayes Approached in Color Detection using Pandas & OpenCV International Journal of Scientific & Engineering Research Volume 11, Issue 12, December- 2020 ISSN 2229-5518.
- [2] Combining Hand Detection and Gesture Recognition Algorithms for Minimizing Computational Cost DOI: November 07, 2020.
- [3] Multiple Object Detection using OpenCV on an Embedded Platform Souhail Guennouni Sidi Mohammed Ben Abdellah University School of Science and Technology Signals Systems and Components Laboratory DOI: 2014 IEEE.
- [4] Face Detection and Recognition Using OpenCV Maliha Khan, Sudeshna, Rani Astya, Shaveta Khepra, 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS) .
- [5] Y. Lu, C. Guo, X. Dai and F. -Y. Wang, "Image Captioning on Fine Art Paintings via Virtual Paintings," IEEE 1st International Conference on Digital Twins and Parallel Intelligence (DTPI), 2021, pp. 156- 159, doi:10.1109/DTPI52967 .2021.9540081.
- [6] Nimisha K P and Agnes Jacob "Brief Review of the Recent Trends in Sign Language Recognition," International Conference on Communication and Signal Processing (ICCSP) Year: 2020, Conference Paper, Publisher: IEEE.
- [7] Abhilash S S, Lisho Thomas, Naveen Wilson, Chaithanya C, "Virtual Mouse using Hand Gesture," International Research Journal of Engineering and Technology (IRJET), p-ISSN: 2395-0072, e-ISSN: 2395- 0056, Volume: 05 Issue: 04 , Apr-2018.
- [8] Ayush Tripath, Arnab Kumar Mondal, Lalan Kumar and Prathosh Ap, "SCLAiR: Supervised Contrastive Learning for User and Device Independent Air writing Recognition," IEEE Sensors Letter Year: 2021, Early Access Article, Publisher: IEEE
- [9] Gangadhara Rao Kommu, "An Efficient Tool For Online Teaching Using OpenCV," 2021 International Journal of Creative Research Thoughts (IJCRT), Volume 9, Issue 6 June 2021, ISSN: 2320-2882
- [10] Yash Patil, Mihir Paun, Deep Paun, Karunesh Singh, Vishal Kisan Borate, "Virtual Painting with OpenCV using Python," First International Conference on Computer Engineering International Journal of Scientific Research in Science and Technology Print ISSN: 2395-6011 , Online ISSN: 2395-602X Volume 5 Issue 8, November-December 2020

