IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Hand Gesture Recognition and Volume Control

Nidhishree Arun[1], Namratha V[2], Ananya Dutta[3], Shreenivas B[4]

Dept. of Electronics and Telecommunication Engineering
B.M.S. College of Engineering
Bangalore, India

Abstract: *A systems that can recognize a hand motion in real time video is hand gesture recognition. Hand gestures are categorized according to their subject matter. The design of hand gesture recognition is one of the more difficult jobs, as it combines two significant issues. The detection of the hand is the first step and creating a sign that can only be utilized by one hand at a time. It can be used in a variety of settings, including human-computer interaction and sign language. The basic concept of hand segmentation and the hand detection system, which use the Haar-cascade classifier, may be used to construct hand gesture recognition using Python and OpenCV. This paper discuss a way for hand gesture identification based on shape-based features detection. The configuration comprises a single camera that captures the user's gesture and feeds it into the system. A fundamental goal of gesture recognition is to develop a system that can recognize specific human gestures and utilize them to send information for device control. With real-time gesture recognition, a user can operate a computer by making a specific gesture in front of a computer's camera. With the help of the OpenCV module, we will create a hand gesture. Without the use of a keyboard or mouse, the system can be controlled via hand gestures.*

Keywords: *Human Computer Interaction, Structuring Elements, Hand gesture, Region of Interest*

1. Introduction

Hand gesture recognition is a system that can recognize a hand motion in a real-time video in everyday life. Hand gestures are classified according to their subject of interest. One of the goal in this implementation is to design a system for hand gesture recognition. One of the most critical issues in computer vision is the task of recognizing hand motions. Human computer interaction (HCI) systems that entail hand processing tasks such as hand detection and hand gesture recognition have become more advanced as information and media technology has progressed.

Detecting and locating the hand in real-time footage from the webcam is the initial stage in any hand processing system. Because of the variety in position, orientation, placement and scale, detecting a hand might be difficult. Variability is also aided by the varying levels of light in the room. Hand gesture recognition often requires numerous levels of processing, including image acquisition, pre-processing, feature extraction, and gesture identification. Using a webcam, image acquisition entails recording an image in a video frame by frame. The collected images are subjected to colour filtering, smoothing as part of the image pre-processing.

Feature extraction is a technique for extracting features from a hand image, such as hand outlines, whereas gesture recognition is a technique for extracting features and recognizing hand gestures. Designing a hand gesture recognition system is a difficult task that comprises two major issues. The first is the detection of a person's hand.

A webcam is used to detect the user's hand in real-time video in this acquisition some issue arises such as inconsistencies in brightness, noise, resolution, and contrast. To identify the gestures, the technique involves segmentation and edge detection and with help of the openCV module we obtain the hand gesture and able to control the volume

2. Literature Survey

Hand gesture detection and utilizing them to control certain set of devices operations and allowing interaction with computer system without the aid of mouse and keyboard. In this paper we draws along the same line but we attributed the use of Haar-cascade classifier to identify hand gesture. Some of the related works in this field are described briefly as follows

[1]A non-local algorithm for hand gestures was proposed by A. Buades, B. Coll, and J. Morel. At the moment, finding finger movement algorithms remains a valid task. Functional analysis and statistics collide. Despite the fact that most recently presented approaches have a high level of sophistication, Algorithms have not yet reached a satisfactory degree of performance applicability. All work admirably when the model matches the algorithm assumptions, but they all fail in general, producing defects in analysing the pixels through the camera. The primary goal of this study is to define a generic mathematical and experimental technique for comparing and classifying conventional hand movement recognition algorithms.

[2] For the no required elements in the video frame, Golam Moktader Daiyan et al. (2012) suggested a high performance decision based median filter. This technique detects noise pixels iteratively over numerous phases before replacing them with the median value. Noise detection is accomplished by enlarging the field of view. Mask till 77% to keep the extraction of local data going. Furthermore, if the algorithm fails to find a noise-free pixel at 7 7, the processing pixel is replaced by the last processed pixel. If the noise-free median value isn't available in the 7th processing window, the last processed pixel is used to determine if it is noise-free. The method chooses a window size if the last processed pixel is noisy. Calculate the number of 0s and 255s in the processing window using the 1515 dimension. Then, in the selected window, replace the processed pixel with 0 or 255, whichever is higher in number.

[3]Rajeshwari Kumar Dewangan et.al accurate object information and obtain a location using a deep learning object recognition technique.Object recognition algorithms are designed based on theSingle Shot MultiBox Detector (SSD) structure, an object recognition deep learning model, to detect objects using a camera.

[4] H. Jabnoun et, al suggested the system that restores a central function of the visual system which is the identification of surrounding objects which is based on the local features extraction concept. Using SFIT algorithm and keypoints matching showed good accuracy for detecting objects.

[5] Košale U, Žnidaršic P, Stopar K suggested that Detection of obstacles is performed by Time of Flight (ToF) sensors, whose ranging data is then processed with an on-board microcontroller and send via Bluetooth connection to the belt. The belt is equipped with a second microcontroller, which interprets the data and presents it to the wearer with 15 vibration motors arranged in a square grid. The glasses are worn on the head, whereas the belt is attached around the stomach area.
But the number of sensors detecting the obstacle decreased with the distance. Circle and square were detected better than triangle. This suggests that different shapes trigger different responses of sensors on glasses.

A. Jaiswal et al. [6] proposed an approach that used user generated picture denoising. The remaining task is broken down into four steps. The first image is denoised using a filtering process, and the second image is denoised using a different method. Wavelet-based approaches are used to denoise pictures, filtering, third hard thresholding, and thresholding. Finally, the approach was applied to a noisy image concurrently. PSNR output results are calculated by comparing all cases, the MSE (mean square error) is obtained. On the basis of PSNR, MSE, and image visual quality, experiments are conducted on 512 X 512 noisy images with noise variance of 0.04, output of median filter, Wiener filter, hard thresholding, and hard thresholding with median filtering. When the filtering and wavelet thresholding techniques are combined, they produce excellent results.

3. Proposed System

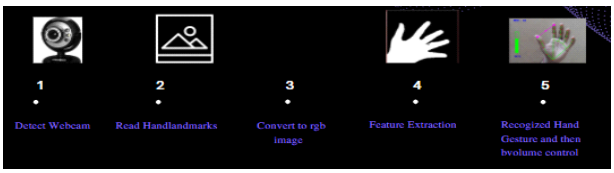

fig. 3.1. Proposed architecture of the system

As shown in figure 3.1, which indicates the proposed architecture of the system used in the volume controller. Here our input is hand gesture which is captured using web camera. Then the GUI (graphical user interface) helps to display the hand that conveys information and it processes the actions of the hand gestures of the user. By recognizing the gestures, the user is able to control the volume of the system which is final output.

4. Methodology

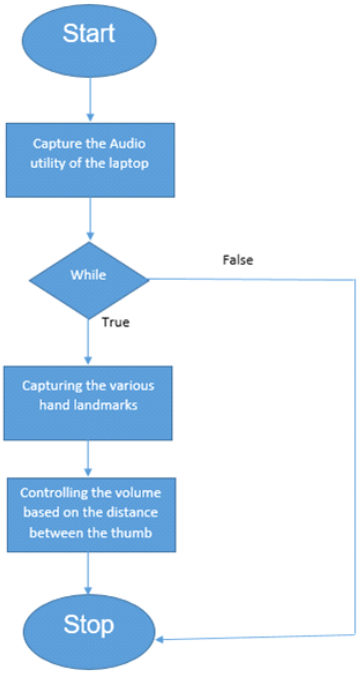

fig. 4.1. Flow chart of the system

The fig 4.1 indicates the flow chart of operation. Firstly, initiate start the program and then import various modules used for the AI recognition and various audio utilities which are of main concern. Next, it capture the area of interest by detecting various contours. Later it execute the loop to detect various hand landmarks. After getting the hand land marks, it verify the distance between the thumb and index figure tip. The frame is displayed giving the final values of the reading with complete decrease and increase in the volume using AI. The program is executed till the loop is iterated. Once it completes the iterations it comes out of the loop and the program stops.

4.1 Algorithm

Step 1: Start the Program
Step 2: Importing the Various Modules-Open Vision used for the AI recognition and various audio utilities which are of main concern.
Step 3: Capturing the Area of Interest by detecting the various contours and differentiating the white and balck region of the interest
Step 4: Executing the Loop to detect the various hand landmarks
Step 5: Getting the Hand Landmarks and verifying the distance between the index and thumb finger based on the algorithm given
Step 6: Display the frame giving the final values of the reading with complete decrease and increase in the volume using Artificial Intelligence. The Program is executed till the loop is iterated once it completes the iterations it comes out of the loop and program stops
Step 7: Stop

5. Implementation

5.1 Software Implementation

Its implemented using the IDE software interpreter Pycharm it can also be implemented using the command prompt as well.

Import the open CV Library to python project which is used as a computer vision tool and to read the image which is nothing but hand in this context. Then we have to use MediaPipe which is a cross-platform framework for building multimodal applied machine learning pipelines. It is used for detection purpose. hypot() method returns the Euclidean norm. The Euclidean norm is the distance from the origin to the coordinates given.

Then to Get default audio device using PyCAW we have used comtypes which is a basic library and audio utilities. The video capture object to capture the information if the video cam is open. Then Media Pipe Hands is a high-fidelity hand and finger tracking solution. If the hands are detected then we have drawn the following outline of the hand using the audio utility function.

Obtain the default audio device using PyCAW.After which we have interfaced to the required volume and then found the range which is from 0 to 100, read the frames from the webcam and convert the image to RGB.

If Fit List of lm or glm Objects with a Common Model is null then we detect hands in the frame with the help of "hands.process()" function. Once the hands get detected we will locate the key points and then we highlight the dots in the key points using cv2.circle, and connect the key points using mpDraw.draw_landmarks. The tip of the index and middle finger then we print(x1, y1, x2, y2).Then we check which fingers are up and we print the fingers. Convert Coordinates and then smoothen the values.

Both Index and middle fingers are close then we reduce the volume and if the index and middle finger are away then we increase the volume. We then find the length of the line through the coordinates. Map the distance of thumb tip and index finger tip with volume range. For our case, the distance between thumb tip and index finger tip was within the range of 15 – 220 and the volume range was from -63.5 – 0.0. Suppose in this case we have taken -8.0 to 194.83366 as maximum volume and similarly the reading for various volumes is taken.

5. Result

In this paper we mainly focused on detecting the hand landmarks and video cam. After which we can access the audio functions inside the system and based on the distance between the thumb and the index finger there will be alteration in the volume.

The figure below shows the average volume detection when the thumb and index finger are in minimal distance from each other.

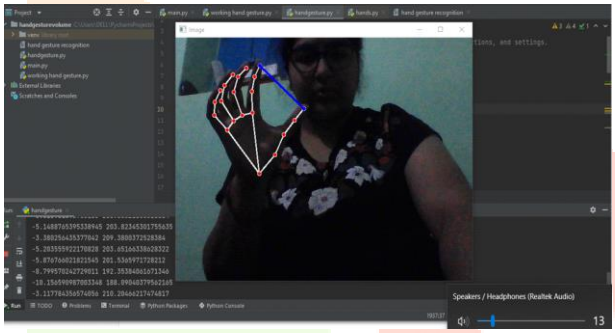

fig. 5.1. Minimum volume

The figure below shows the average volume detection when the thumb and index finger are in minimal distance from each other.

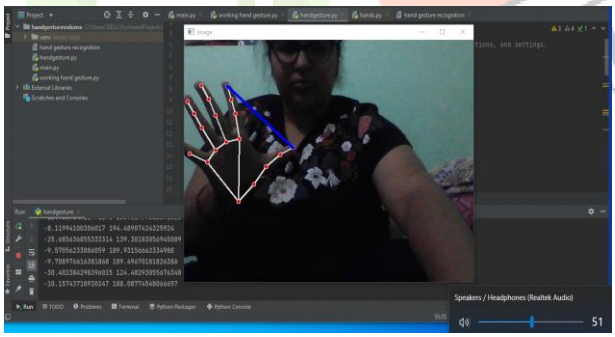

fig. 5.2. Moderate volume

The figure below shows the maximum volume detection when the thumb and index finger are away from each other.

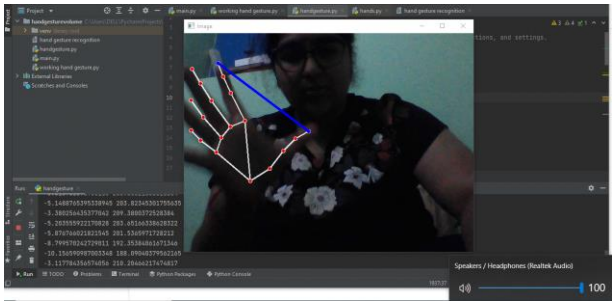

fig. 5.3. Maximum volume

6. Conclusion

In this paper we have taken up is a vision-based hand Gesture system that runs on a code based on Open CV library of python. It makes used of various algorithms and methods such as tracing significant points in the images and distance calculation between points. Specifically, the system can track the tip positions of the counters and index finger for each hand. It is an efficient and simple way to handle sound devices without much manual work. It does not require any special markers or gloves and can operate in real-time on a commodity PC with low-cost cameras.

ACKNOWLEDGMENT

The authors acknowledge the support and encouragement of the management of B.M.S. College Engineering, Bengaluru.

REFERENCES

[1] Coll, Bartomeu, Antoni Buades & Morel, Jean-Michel. (2005). A Review of Image Denoising Algorithms, with a New One. SIAM Journal on Multiscale Modeling and Simulation. 4. 10.1137/040616024.

[2] Daiyan, Golam & Mottalib, M.A.. (2012). High performance decision based median filter for salt and pepper noise removal in images. 107-112. 10.1109/ICCITechn.2012.6509707.

[3]Rajeshwari Kumar Dewangan, Siddharth Chaubey, "Object Detection System with Voice Output using Python," *2021 International Journal of Science & Engineering Development Research*, Vol. 6, Issue 3, pp. 15-20, 2021.

[4] H. Jabnoun, F. Benzarti, H. Amiri, "Object detection and identification for blind people in video scene," *2015 15th International Conference on Intelligent Systems De- sign and Applications (ISDA)*, pp. 363-367, 2015.

[5] Košale U, Žnidaršic P, Stopar K, "Detection of different shapes and materials by glasses for blind and visually impaired," *2019 6th Student Computer Science Research Conference*, Vol. 57, 2019.

[6]Ayushi Jaiswal Ajay Somkuwar Jayprakash Upadhyay '*Image denoising and quality measurements by using filtering and wavelet based techniques'* AEU - International Journal of Electronics and Communications 68(8):699-705

[7]E. Barri, A. Gkamas, E. Michos , C. Bouras, C. Koulouri, S.A.K. Salgado, "Text to Speech through Bluetooth for People with Special Needs Navigation," *International Conference on Ne8tworking and Services*, 2010.

[8] H. A. Karimi, M.B. Dias, J. Pearlman, G. J Zimmerman, "Wayfinding and Navigation for People with Disabilities Using Social Navigation Networks," *EAI Endorsed Transactions on Collaborative Computing*, Vol. 1, Issue 2, e5, October, 2014.

[9] Rahman Mohammad Marufur, Md. Milon Islam, Saeed Anwar Khan, Shishir Ahmmed "Obstacle and fall detection to guide the visually impaired people with real time monitoring," SN Computer Science, Vol. 1, pp 1-10 (219), 2020.