JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

HAND GESTURE RECOGNITION USING HAAR CASCADE ALGORITHM IN MEDIA PIPE FOR SIGN LANGUAGE AND GESTURE CONTROL

Dr. J.Arokia Renjith, M.R.U.Preethika, M.S.Sai Prasanthini, Sokitha K
Head of the computer science and engineering department, Student, Student, Student
Computer Science and Engineering, Jeppiaar Engineering College, Chennai, India

Abstract **:** Recognizing hand gestures from an image has been significant for the future technology into creating new systems and coming up with solutions for advanced level of human and computer interactions. In this work, we have created a system which would aid to the future technology in the human-computer interactions. Our work aids in two ways, one is to provide a sign language interpretation using hand gestures, another is to use gestures in controlling the computer system partially. We have come up with our very own sign language which will be helpful for visual communication with the deaf and mute people. And in this work we have also come up with a novel way of partially controlling a computer screen such as controlling a video, a webpage etc., with the use of hand gesture. This system is vision based and uses machine learning algorithm and inputting the images through a computer web camera. Initially we would get the images through the web camera, then the hand region is extracted using background subtraction, and then fingers are segmented. And finally a rule classifier is applied to detect the hand gesture. We have used OpenCV and MediaPipe.

IndexTerms **- Hand Gesture, Open CV, Machine Learning.**

I. INTRODUCTION

In today‘s world, the computers have become an important aspect of life and are used in various fields however, the systems and methods that we use to interact with computers are outdated and have various issues. Usually the basic way in which human interact with computers is by using basic pointing device, keyboard, touch pad or an advanced voice recognition system. Dumb people are usually deprived of normal communication with other people in the society. Also normal people find it difficult to understand and communicate with them. As a normal person is not aware of the various gestures that are part of a sign language so the sign language gets limited such that only the deaf and mute people and their family could understand. Hence In our proposed project, the hand gesture recognition system with the help of programming using python, opencv and Mediapipe to facilitate interaction with the computer through the computer web camera. Initially we use background subtraction and HSV segmentation to create a mask, after which the hand is segmented, with which we will find the convex hull and convexity hull to detect the number of fingers raised and the gesture shown. After identifying the gesture the sign language interpretation as well as the controlling the computer screen can be done.

1.1 Scope Of The Project

A solution of communication with the deaf-mute people is done by using the services of sign language interpreter. But usually the usage of sign language interpreter is not cost efficient. Cost-effective solution is required so that the deaf-mute and normal people can communicate normally and easily. Also improving the technology for human and computer interaction, rather than the basic devices which can restrain the user and could limit the user if the device gets dropped or broken. We can improve this with controlling the screen with just hand gestures in the mid-air. Our strategy involves implementing such an application which detects pre-defined hand gestures for sign language as well as to control the system. For detecting the gesture, we use basic level of hardware component like an inbuilt camera in a laptop or a web camera. Our application would be a comprehensive User-friendly Based system built on OpenCV module. Instead of using

technology like gloves or Kinect, we are trying to solve this problem using state of the art computer vision and machine learning algorithms. This application simply detects the gesture and displays appropriate alphabet and the control.

1.2 Motivation Of The Work

We had chosen this project with an interest of creating a direct interaction of humans with the electronic devices. This makes the user to experience a new level of interaction. The gesture control technology would reduce our dependence on the devices such as mouse, keyboard etc., and also it would reduce the complexity of controlling a system. Initially this technology is widely used in gaming (like Xbox Kinect), but the application of motion/gesture control technology would be more diverse if we apply it to day to day used electronics such as computers, televisions, etc., for various gesture control such as clicking, scrolling etc. And also we can create our very own sign language, which would provide an easy way of interaction with the deaf and mute community as sign language interpreters are not cost efficient.

1.2.1 Machine Learning

Machine learning is an application of AI that gives systems the power to automatically learn by themselves and improve from experience without the need of to be programmed. Machine learning focuses on the event of computer programs that may access data and use it learn for themselves. The method of learning begins with observations or data, like examples, direct experience, or instruction, so as to seem for patterns in data and make better decisions within the future supported the examples that we offer. The first aim is to permit the computer to learn automatically without human intervention and determine actions accordingly.

1.3 Problem Description

There are generally two approaches for hand gesture recognition, which are hardware based, where the user must wear a device, and the other is vision based which uses Machine learning techniques with inputs from a camera. The proposed system is vision based, which uses machine learning techniques and inputs from a computer webcam. The input frame would be captured from the webcam and systems are generally broken down into five stages, data obtaining (capturing the image), data pre-processing (skin detection and background subtraction), Feature Extraction (hand contour extraction), assigning the gesture and Predicting the gesture. The hand contour is found and used for hand tracking and gesture recognition. With the gesture recognized we can control the computer screen by selecting, scrolling and used in video controls and also used as a sign language interpreter. The Flowchart of the project is as follows:

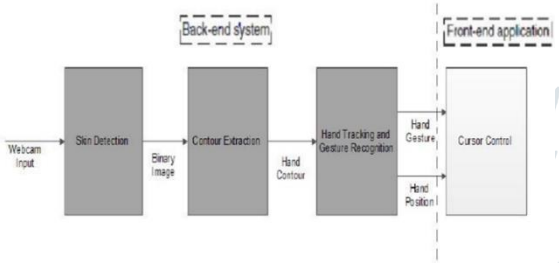

Figure 1.3.1 Flowchart

II LITERATURE SURVEY

2.1 Sign Language Recognition Using Multiple Kernel Learning: A case study of Pakistan Sign Language, Farman Shah; Muhammad Saqlain Shah; Waseem Akram; Awais Manzoor; Rasha Orban Mahmous; Diaa Salama Abdelminaam, IEEE Acess,2021

All over the planet, deaf individuals use signing because the solely reliable supply of communication with one another similarly like traditional individuals. These human action signs square measure created of the form of the hand and movement. In Asian country, deaf individuals use Asian country signing (PSL) as a way of communication with individuals. In scientific literature, several studies are done on PSL recognition and classification. Most of those work targeted on color-based hands whereas some others square measure sensors and Kinect-based approaches. These techniques square measure pricey and additionally avoid user-friendliness. During this paper, a method is projected for the popularity of xxxvi static alphabets of PSL victimisation vacant hands. The dataset is obtained from the signing videos. At a later step, four vision-based options square measure extracted i.e. native binary patterns, a bar graph of homeward-bound gradients, edge-oriented bar graph, and sped up sturdy options. The extracted options square measure one by one classified victimisation Multiple kernel learning (MKL) in support vector machine (SVM). We tend to utilize a one-to-all approach for the implementation of basic binary SVM into the multi-class SVM. A pick theme is adopted for the ultimate recognition of PSL. The performance of the projected technique is measured in terms of accuracy, precision, recall, and F-score. The simulation results square measure promising as compared with existing approaches.

2.2 Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues, Muhammad Al-Qurishi; Thariq Khalid; Riad Souissi, IEEE Access, 2021

People with hearing impairments area unit found worldwide; so, the event of effective native level linguistic communication recognition (SLR) tools is important. we tend to conducted a comprehensive review of machine-controlled linguistic communication recognition supported machine/deep learning strategies and techniques printed between 2014 and 2021 and over that the present strategies need abstract classification to interpret all offered knowledge properly. Thus, we tend to turn our attention to parts that area unit common to the majority linguistic communication recognition methodologies. This paper discusses their relative strengths and weaknesses, and that we propose a general framework for researchers. This study additionally indicates that input modalities bear nice significance during this field; it seems that recognition supported a mix of information sources, together with vision-based and sensor-based channels, is superior to a uni-modal analysis. Additionally, recent advances have allowed researchers to manoeuvre from straightforward recognition of linguistic communication characters and words towards the capability to translate continuous linguistic communication communication with tokenish delay. However, the pace of analysis is encouraging, and any progress is anticipated if specific difficulties area unit resolved.

2.3 Hand Gesture Recognition System for Touch-Less Car Interface Using Multiclass Support Vector Machine Authors: Mrinalini Pramod Tarvekar, 2018

Touch-less automotive interfaces are becoming the eye in automobile industries. By victimisation the hand gestures its potential to regulate some activities of cars, for this effective recognition system for hand gesture is needed. This paper presents the popularity system of hand gesture for touch-less automotive interface. This technique accepts video frames sequence and segmentation is applies on theses frames. Here the new segmentation technique is applied by detective work skin portion victimisation HSV, YCgCr and YCbCr color area. From the segmental pictures their color and edge characteristics square measure extracted and so hold on within the information with their various labels. Edge bar graph descriptor is employed for retrieving the form characteristics from pictures wherever a color structure descriptor (CSD) is employed to capture the spatial distribution of colour in a picture. Then for identification of gesture multiclass SVM classifier is employed. For the experiment, hand gesture information from Cambridge is employed wherever video based mostly frames of pictures square measure gift for nine gestures.

2.4 Performance analysis of RTEPI method for real time hand gesture recognition Authors: A.V. Dehankar ; Sanjeev Jain ; V. M. Thakare, 2017

Hand gestures offer a natural and intuitive thanks to move with the computers, mobile devices, etc. several researchers ar performing on gesture recognition wherever the aim is to spot and distinguish human gestures so known gestures ar wont to management applications in specific domains. Recognizing real time hand gestures is incredibly difficult and troublesome task. This paper presents a unique Real Time finish purpose Identification technique from a true time video of hand gesture mistreatment pc vision and Image process techniques the $64000 Time finish purpose Identification (RTEPI) technique mentioned in [1] relies on correct finish purpose Identification [2], that permits the AEPI technique to figure on real time hand gestures captured through internet camera or laptop computer camera. The RTEPI technique provides the proper frame for real time process to AEPI technique that then detects the correct technique. The AEPI technique has been enforced to deal with the issues of varied background, luminance, blurring etc. 5 totally different phases of AEPI technique includes pre-processing, centre of mass detection, removal of unwanted objects, cutting and recognition that ar already mentioned in [2,3]. The paper presents the result and performance analysis of RTEPI technique for all doable input patterns of real time hand gesture recognition.

2.5 A Robust Hand Gesture Recognition Method via Convolutional Neural Network Authors: Xing Yingxin ; Li Jinghua ; Wang Lichun ; Kong Dehui, 2016

Hand gesture plays a crucial role in nonverbal communication and natural human-computer interaction. However, the advanced hand gesture structure and numerous surroundings factors result in low recognition rate. For example, hand gesture depends on people, and totally different individuals' hands area unit with different sizes and postures, additionally, free environmental illumination conjointly influences hand gesture recognition performance. Therefore, hand gesture recognition remains a difficult issue. This paper proposes a strong methodology for hand gesture recognition supported convolutional neural network that is employed to mechanically extract the spatial and linguistics feature of hand gesture. Our methodology consists of a changed Convolutional Neural Network structure and information pre-processing, that corporately increase hand gesture recognition performance. The experimental results on each Cambridge Hand Gesture Dataset and self-constructed dataset show that the projected methodology is effective and competitive.

III. SYSTEM DESIGN

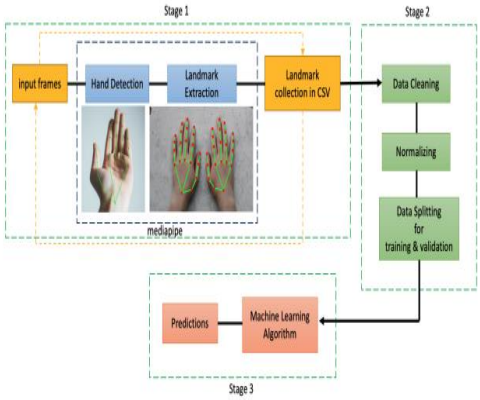

Figure 3.1 System Architecture

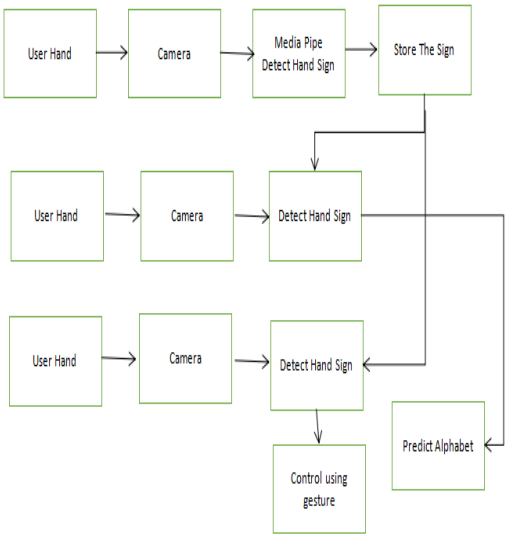

Figure 3.2 System Flowchart

IV. SYSTEM IMPLEMENTATION

The main scope is to detect the hand gesture from an image or a live video and then identify the gesture and perform the certain functionality such as sign language or to interact with the computer system. The proposed method is obtaining the image through data pre-processing, measuring the entropy, separating hand region from images, tracking the hand region and recognizing hand gestures. The convexity defects for the hand contour was calculated using the Open CV inbuilt function "CV Convexity Defects". Using the Mediapipe hand landmark, we would find the position of the hand in the image, then create a sequence of array to store the data's of the position of the fingers. This is then compared to the dataset and performed.

Step-1: Input video stream
Step-2:Pre-processing the stream for background modelling
Step-3: Foreground detection
Step-4: Obtain the foreground mask
Step-5: Background modelling
Step-6: Finger detection
Step-7: Perform appropriate action

Advantages:

1 Fast and reliable for recognition system and powerful results from proposed algorithm.
2 Good performance system with complex background.
3 Automatic sampling, and augmented filtering of the data improved the performance.
4 Accurate shape of the hand obtained led to good feature extraction and recognition.

4.1 HAAR Cascade Algorithm

We have used HAAR Cascade algorithm as this algorithm is very efficient in background subtraction. HAAR Cascade is an object detection algorithm in machine learning which is used to identify objects in an image or a real time video. HAAR like feature based classifier provides both high accuracy & speed. It needs less instructions and has less

false detections. Use of integral images causes high speed of analysis while rectangular property of the HAAR like features characterize non symmetrical properties of Gesture appearance, so it is perfect for Gesture detection procedure.

The algorithm has four stages:

1. HAAR Feature Selection
2. Creating Integral Images
3. Ada boost Training
4. Cascading Classifiers

Advantages:

1 We can able to train more amount of data
2 High accuracy

4.2 Modules Description

4.2.1 Data capturing

The initial move is to capture the image from camera and to define a region of Interest in the frame, it is important as the image can contain a lot of variables and these variables can result in unwanted results and the data that needs to be processed is reduced to a large extent. To capture the image a web-camera is used that continuously captures frames and is used to get the raw data for processing. The input picture we have obtained is uint8. The Procured image is RGB and must be processed before.

4.2.2 Data Pre-Processing

Pre-processing method is done in a 2-steps process:

4.2.2.1 Segmentation

First Process is the Segmentation process. We change rgb to grey-scale picture then into the binary picture. Background subtraction was then performed to remove the hand and other skin color objects in the background so we can have just two Area of Interest in picture. That is, one will be the hand (foreground) and another one is background. Algorithm used is HAAR Cascade for this process and grey scale picture are converted into binary picture having an area of interest as the hand and the background.

4.2.2.2 Morphological filtering

After thresholding we have to make sure that there will be no noise is present in image, so we are using morphological filtering Techniques.

4.2.3 Feature extraction

Pre-prepared or pre-processed picture is accessible to be utilized and different highlights of the resultant picture are removed. Following are the features which are extracted:

- Finding Contours – by using openCV
- Finding and correcting convex hull – by using openCV
- Action – the gesture shown

We use OpenCV for finding the convexity defects (the convex hull is the outline of a shape, contour is the boundary of the shape) and then using media pipe we extract the hand and find the hand landmarks, fingers position. With the use of mediapipe we draw the skeleton or the landmarks on the hand.

4.2.4 Assign Gesture

After Extracting the hand, with the help of the stored data we assign the gesture they are in the format as [0,0,0,0,0] to [1,1,1,1,1] in an array , while 0 represents if the finger is not raised and 1 represents if the finger is raised. After which we assign the given images an alphabet for each and unique gesture in the case of sign language, and then each gesture would be allocated to a function such as scrolling, selecting etc.

4.2.5 Prediction

After assigning alphabet to each gesture and after assigning to a function to control the computer screen now we can able to predict from the gesture given by the user. First we need to specify whether to control the system or use sign recognition. After specifying we can perform the gesture and then we can predict from the gesture given by the user.

V. TESTING

5.1 Software testing

In a generalized way, we can say that the system testing is a type of testing in which the main aim is to make sure that system performs efficiently and seamlessly. The process of testing is applied to a program with the main aim to discover an unprecedented error, an error which otherwise could have damaged the future of the software. Test cases which brings up a high possibility of discovering and error is considered successful. This successful test helps to answer the unknown errors.

VI. CONCLUSION

The hand gesture recognition system for sign language interpretation and for controlling computer screen was developed in the Python language, using the Open CV library and with Mediapipe Framework. The system was able to control the screen with hand gesture shown as well as identify the sign for interpretation. The system has the potential of being a viable replacement for the computer controlling such as mouse or a touchpad or a keyboard, however due to the constraints encountered; it cannot completely replace them. The major constraint of the system is that it must be operated in a well-lit room, since it is very common for computers to be used in outdoor environments with poor lighting condition. What we do is, we just open our script file it will automatically launch the camera. After camera is launched we can choose whether to control the screen or to use the sign language interpretation. Then system invokes the tools that we required to run it for instance- Open CV, Mediapipe, Camera, pyautogui. Now, we are ready to do just sit back and control without using any conventional method.

VII. FUTURE ENHANCEMENT

We would improve the performance of the software especially to completely use only gestures to control the system and using hand tracking as a replacement for computer mouse. We would also like to extend the controlling of the system to a general purpose computer controller as to completely remove the usage of mouse or keyboard for controlling the computer. We would also like to improve the sign language interpretation system as to allow a complete interpretation allowing both the parties (the deaf and mute and a normal person) to communicate with ease, with adding voice translation.

REFRENCES

[1] Sign Language Recognition Using Multiple Kernel Learning: A Case Study of Pakistan Sign Language, Farman Shah; Muhammad Saqlain Shah; Waseem Akram; Awais Manzoor; Rasha Orban Mahmoud; Diaa Salama Abdelminaam. IEEE Access, 2021 **https://ieeexplore.ieee.org/document/9422714**

[2] Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues, Muhammad Al-Qurishi; Thariq Khalid;Riad Souissi. IEEE Access, 2021 https://ieeexplore.ieee.org/document/9530569

[3] Hand Gesture Recognition System for Touch-Less Car Interface Using Multiclass Support Vector Machine, Mrinalini Pramod Tarvekar, 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS) https://ieeexplore.ieee.org/document/8663003

[4] Performance analysis of RTEPI method for real time hand gesture recognition, A.V. Dehankar ; Sanjeev Jain ; V. M. Thakare, 2017 International Conference on Intelligent Sustainable Systems (ICISS) https://ieeexplore.ieee.org/document/8389335

[5] A Robust Hand Gesture Recognition Method via Convolutional Neural Network, Xing Yingxin ; LiJinghua ; Wang Lichun ; Kong Dehui, 2016 6th International Conference on Digital Home (ICDH) https://ieeexplore.ieee.org/document/8039082