International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9, Issue-7, May 2020

Hand Gesture Communication for Blind Communication using Raspberry Pi

Punit Dobriyal, Diparna Adhikary, Abhimanyu Jain, G. Rajkumar

Abstract: ***HRI represents an obstacle to represent how humans and robots interact. The intricacy is that a robot doesn't comprehend the human language straightforwardly and HRI requires media for correspondence which can be comprehended by the robot and effortlessly done by human, especially to help old and deaf people, recovering patients, in this manner signal acknowledgment as correspondence media is expected to provide a request to the robot. Machine learning is a fragment of Artificial Intelligence (AI) which discusses the development of a system that depends on information or data. This paper enunciates about the hand signals as input for Bioloid Premium Robot utilizing two strategies, Fuzzy C Means clustering and Support Vector Machine (SVM) with directed acyclic graph (DAG). Here, decision K-Means clustering or Lloyd's algorithm suggested the way to clustering some data by using the Euclidean idea of distance between all the present data elements.***

Keywords: Clustering, Image recognition, Processing, Digital Processing, Robotics, AI

I. INTRODUCTION

The Hand gesture recognition system is becoming more common as time passes by. This system is exclusively used by people with some sort of an impairment. The use of this system has increased as the awareness among people has increased as well. The principle point of this framework is to give an approach to control electronic apparatuses utilizing hand signals. The applied framework here is related to an efficient method of detecting the hand gestures or the hand signals, without utilizing any of the external devices such as sensors or the manual method of using a glove. The system can recognize the hand signals using the webcam and a couple of algorithms. The astute piece is to decipher assortments of these shapes as single items, for example autos on a street, boxes on a transport line or destructive cells on a magnifying lens slide. One explanation that this is a simulated intelligence issue, is that an item can show up totally different when seen from various points or under various lighting. Another issue is choosing what highlights have a place with what item and which are foundation or shadows and so on.

Revised Manuscript Received on May 30, 2020.
***** Correspondence Author
Punit Dobriyal*, Department of Computer Science & Engineering from SRM Institute of Science & Technology, Chennai, India.
Diparna Adhikary, Department of Computer Science & Engineering from SRM Institute of Science & Technology, Chennai, India.
Abhimanyu Jain, Department of Computer Science & Engineering from SRM Institute of Science & Technology, Chennai, India.
Mr. G Rajkumar, Assistant Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India.

The human visual framework plays out these assignments for the greater part unknowingly, however a PC requires dexterous programming and heaps of preparing capacity to move toward human execution. Controlling information as a picture through a few potential systems. A picture is normally deciphered as a two-dimensional cluster of brilliance esteems, and is most naturally spoken to by such examples as those of a photographic print, slide, TV screen, or film screen. A picture can be handled optically or carefully with a PC.

For digitally processing an image, it is first important to reduce the image to a series of bits that can be read by the computer. Each bit represents the brightness magnitude of the image at a specific location and is referred to as a pixel. These activities, taken singly or in blend, are the methods by which the picture is upgraded, re-established, or compacted. The procedure of recognizing the important details in real-world images is one of the major aspects in Computer vision. In general, contortions from foundation mess, scale, and perspective varieties can render appearances of even a similar article case to be inconceivably unique.

Thus, models for object classes must be adaptable enough to suit class changeability, yet discriminative enough to sifter out evident article cases in jumbled pictures. This paper works essentially on two objectives, that is, picture grouping and item discovery. The assignment of picture grouping is to decide whether an article class is available in a picture, while object location restricts all occurrences of that class from a picture. In general, there are 3 types of images used in Digital Image Processing. They are Binary image, Gray Scale Image and Color Image.

There are 2 algorithms used here mostly - K-Means Clustering and Support Vector Machine.

The algorithm, K-means clustering is a form of autonomous or unbridled learning, which is used when the user has unlabeled data. Unlabeled data refers to a set of data or collection of data which is not defined or is not under a certain class.

The main purpose of this calculation is to look for bunches in the information, with the quantity of clusters spoken to by the variable K. The calculation works repeatedly to allot every datum point to one of the K groups depending on the highlights that are given. Information focused are grouped depending on and include closeness. The κ-implies grouping calculation utilizes iterative refinement to deliver a conclusive outcome. The calculation inputs are the quantity of clusters κ and the informational collection. The informational index is an assortment of highlights for every datum point. The calculations begin with introductory evaluations for the κ centroids, which can either be arbitrarily created or haphazardly chosen from the informational collection.

The paper is based on python and uses python language for implementation of the project.

We are using OpenCV, which is a library of programming functions mainly focused at real-time computer vision. The library is free to use, that is it can be used by any user. It is cross-platform in nature. So, all the necessary functions can be obtained from OpenCV.

II. PROPOSED SYSTEM

The proposed approach introduces a change in the sorting algorithm used and a couple of new instruments in the already existing system. The already existing approach of recognizing the hand signals is not changed or interfered with, in any way. The new algorithm that is to be used in the proposed system is Fuzzy-C-Means Clustering algorithm, Support Vector Machine and Probabilistic Neural Network Classifier. The existing system uses K means sorting algorithm which is lacking in terms of performance and its ability to handle outliers. Fuzzy-C-Means Clustering algorithm is also called the Soft K-Means method.

The Support Vector Machine is a machine learning model which is supervised in nature. It is a classification algorithm which can work very well with a limited set of data. The model creates a line and splits the data into various classifications, mostly into two types. This line is known as Hyperplane. It is also called Decision Boundary. The Hyperplane is responsible for the classification of data provided to the Support Vector Machine. It can easily manage linearly separable data. The non-linear data is not a problem for the Support vector Machine as well. It can still operate on the dataset using a few tricks such as kernel functions. The Kernel converts the lower dimension data into the required higher dimension data. For example, a 2-D graph is converted into a 3-D graph.

We are using the Fuzzy-C-Means algorithm in the proposed system. Compared to this, the K-means algorithm separates the dataset into non-overlapping subgroups or clusters. The algorithm uses centroids to create the clusters. The Fuzzy-C-Means does not put the restriction that a data node cannot belong to another cluster, that means, a data node can belong to more than one cluster. The K-Means clustering algorithm is not able to handle complicated cluster shapes. The measures used to verify the clusters include distance, connectivity and intensity. We use the K-Means clustering algorithm in image compression as well.

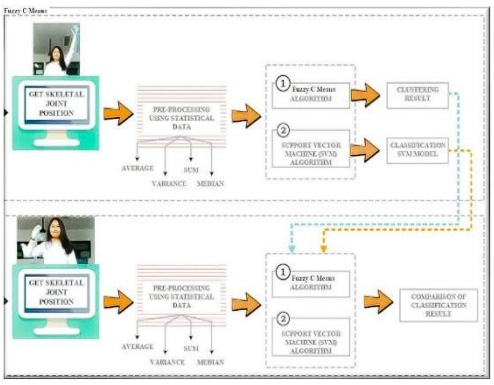

Fig 1: Architecture Diagram

The Probabilistic Neural Network Classifier will help in choosing the correct decision by the system. As the name suggests, it works on Probability Distribution functions. It is used generally in classification and recognition problems. It works on 4 layers namely, Pattern Layer, Input Layer, Summation Layer and Output Layer.

The Input Layer calculates the distance from the input vectors to the training vectors given in the model. The Pattern Layer uses Gaussian functions which are formed using the training dataset. It stores the estimations of the indicator factors for the case alongside the objective worth. A shrouded neuron processes the Euclidean separation of the experiment from the neuron's inside point and afterward applies the RBF portion work utilizing the sigma esteems. The summation layer just summarizes the outputs received from the second layer, that is, pattern layer. Then, the output layer selects the largest value and then the associated class is the solution. The thing to consider here is that the Probabilistic Neural Network is insensitive to outliers. The accuracy of the neural network is considered to be one of the best in the market currently. The system uses the training dataset for creating a model and then the various algorithms are responsible for producing the correct output. In this case, the correct command provided by the user.

III. IMPLEMENTATION

Hand gesture recognition is a really fascinating topic for computer vision enthusiasts to work on. There are various differences between how a human visualizes a scene and how a computer will look at the same scene. Humans have the capacity to perceive any real scene or images and attribute each aspect of the scene to a category.

Fig 2: Sample Image

For instance, in the above image, as a human ourselves, we can perceive the various aspects of the image, like, the background of trees, even the single bigger tree, lot of people (both male and female) walking around in the background, and a woman facing the camera, who also seems to be holding a handbag of some sorts. The level of detail a human mind can perceive is even higher, from the expression of the person facing the camera, to the artistry of the camera work, it can be said to be truly unmatched. While when considering a computer, what it may see might depend on the level of learning the system has had. The concept of controlled and autonomous learning could be applied to this, in a way that, only if a computer knows what a woman looks like, it can see a picture and recognize one.

Fig 3: Image visible to the machine

The same image may be perceived by the computer similar to what the image above has described. The level of detail in the above image is by far inferior to what we humans perceived with the earlier image. Here we can see that the computer has noticed the person standing in the front. In the background, it has noticed the grass in the ground, the trees and the sky. But others may be irrelevant, yet present, people walking in the background, facial expressions of the person, perhaps even the gender, and other things haven't been seen by the computer.

So, we begin the implementation by what is called as the background subtraction. In background subtraction, we teach the computer to neglect the background features of an image or a video, and to focus on what's dominating the whole scene. In the earlier image, this aspect was the woman facing the camera. In the figure below, we can see what the process of background subtraction is all about. What appears to be in focus seems to be a boat. So, when we subtract the background of water, we get the boat alone.

Other techniques used here are contour extraction and motion detection. Motion detection logic can be taught to a system to detect specific hand gestures. Motion detection sensors are used to sense which part of the body is moving, and based on that the gesture would be identified and then compared with the knowledge base to make a choice as to what a person wants.

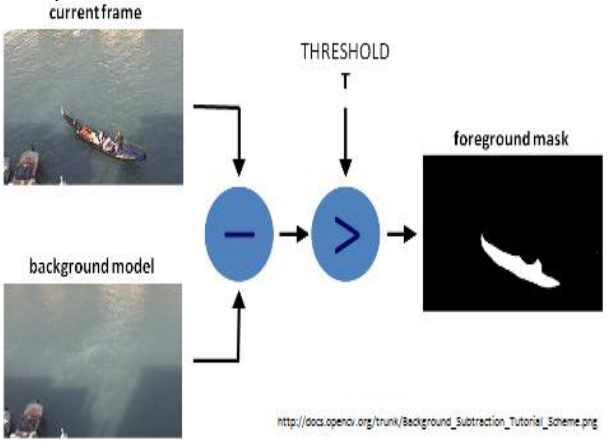

Fig 4: Filtering out background image

The figure below is an example of real time video feed being used to recognize what the person is gesturing. As we can see from the picture below, the hand for instance can be seen with all the five fingers outstretched. Once the motion of stretching the fingers is done, the motion detector will sense the motion of the fingers, and the captured image will be then detected, read and understood.

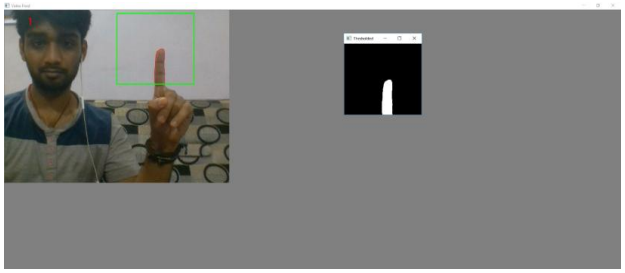

Fig 5: Reading the sample input(1)

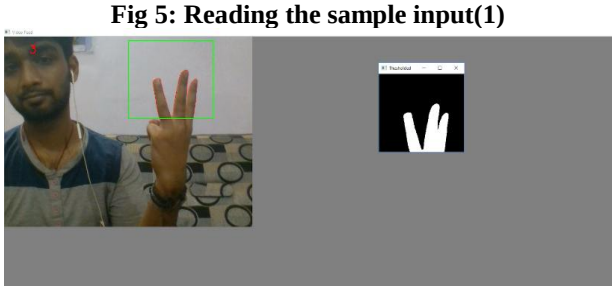

Fig 6: Reading the sample input(2)

The process can be summarized into 5 simple steps –

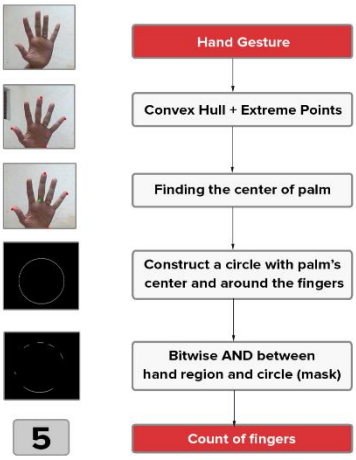

Fig 7: Process

The above image talks about the following steps -

-Firstly, locate the arched frame of the fragmented hand district (which is a form) and figure out the most extraordinary focuses in the raised structure (Extreme Top, Extreme Bottom, Extreme Left, Extreme Right).

-Then, locate the focal point of the palm utilizing these boundaries focused in the curved frame.

-After that by utilizing the palm's middle, develop a hover with the most extreme Euclidean separation (between the palm's inside and the extraordinary focuses) as sweep.

-Lastly, perform bitwise AND activity between the thresholded hand picture (outline) and the round ROI (cover). This uncovers the finger cuts, which could additionally be utilized to calculate the quantity of fingers appeared.

IV. RESULTS AND DISCUSSION

The method of communication implemented using the concept of machine learning was executed successfully. The image was converted to data for the robot in various instances, one as an example is shown in Fig 6.

The objects produced, a human hand here, was recognized by the software correctly and efficiently. Feeding information to the machine will take time for complete functionality of all commands. Increasing the information gathering rate will help for a faster setup of the system.

V. CONCLUSION

Through the collection and analysis of the achieved result via the implementation, it is henceforth, concluded that this process is used for color segmentation and template matching work with high accuracy, when working with hand gesture recognition. The results acquired are appropriate, and can be implemented using a mobile smart phone device with a frontal camera. Not everyone understands sign language, this method of communication helps in working with machines and people alike.

REFERENCES

1. Paulraj M. P. Sazali Yaacob, Mohd Shuhanaz bin Zanar Azalan, Rajkumar Palaniappan, "*A Phoneme based sign language recognition system using skin color segmentation*", *Signal Processing and its Applications* (CSPA) – pp: 1 – 5, 2010.
2. Byong K. Ko and H. S Yang, "*Finger mouse and gesture recognition system as a new human computer interface*", pp: 555-561, 1997
3. Manar Maraqa, Dr. Raed Abu Zaiter, "*Recognition of Arabic sign Language using recurrent neural networks*", *Applications of Digital Information and Web Technologies*, pp: 478 – 481, 2008.
4. Yang quan, "*Chinese Sign Language Recognition Based on Video Swquence Appearance Modeling*", *Recognition using HMM*, pp: 623 – 626, 2006.
5. P. Mekala, R. Salmeron, Jeffery Fan, A Davari, J Tan, "*Occlusion Detection Using Motion-Position Analysis*" ICIEA, the 5th *IEEE Conference*, pp: 1537 – 1542, 2010.
6. K. Kawahigasi, Y. Shirai, J. Miura, N. Shimada "*Automatic Synthesis of training Data for Sign Language*" *IEEE 42nd South-eastern Symposium*, on System Theory (SSST"10), pp: 197-201, 2010.
7. Jae Y. Lee and Suk I. Yoo "*An Elliptical Boundary Model for Skin Color Detection*" pp: 2- 5, 2002.
8. *Digital Image Processing Using Matlab* by Ganzalez, p: 205. 2009.

AUTHORS PROFILE

Punit Dobriyal was born in Chipeldhungi, Uttarakhand, India in 1999. He is expected to complete his Bachelor of Technology (B-Tech) in the branch of Computer Science & Engineering from SRM Institute of Science & Technology in 2020. His research interests include machine learning, game development, cyber security and ethical hacking.

Diparna Adhikary was born in Chirkunda, Jharkhand, India in 1998. He is expected to complete his Bachelor of Technology (B-Tech) in the branch of Computer Science & Engineering from SRM Institute of Science & Technology in 2020. His research interests include, virtual reality, Cyber security and Internet of Things (IoT).

Abhimanyu Jain was born in Udaipur, Rajasthan, India in 1998. He is expected to complete his Bachelor of Technology (B-Tech) in the branch of Computer Science & Engineering from SRM Institute of Science and Technology in 2020. His research interests include Web designing, Marketing and Cyber Security.

Mr. G. Rajkumar born in Tamil Nadu. Completed his Bachelor of Engineering in the branch of Information Technology from CEG, Anna University, Chennai 2008. He completed his Master of Engineering in Information Technology from CEG Anna University, Chennai at 2011. He's working as Assistant Professor in the Department of Computer Science and Engineering at SRM Institute of Science and Technology Chennai from 2012 till date. His research interest includes Wireless Sensor Networks.