ISSN: 0974-5823 Vol. 6 No. 3 December, 2021

International Journal of Mechanical Engineering

A GENERAL SOLUTION TO SIGN LANGUAGE RECOGNITION FOR DIFFERENT SKIN TONES

Krati Omer[1a)] and Dr. Shalini Gupta[2b)]

Author Affiliations

[1]*M.Tech, Kanpur Institute of Technology, Kanpur, Uttar Pradesh, India*

[2]*Kanpur Institute of Technology, Kanpur, Uttar Pradesh, India*

Abstract Deaf-Mute people communicate through their sign language which is well known in their community but not easily understood by other people. Sign language recognition system is very helpful in this area but as in our world people of different skin tones live similar diversity exists in deaf-mute community. Need of SLR for different skin tones arises. This paper proposes the general solution to acquire data which is independent of skin tone. With the combination of ROI, Background Subtraction, Connected Component Algorithm data is acquired. Then model is trained using CNN Architecture. This paper gives the proposal, its real time system should be developed in future.

Keywords : Sign language, skin tone, Background Subtraction, Connected Component, ROI, CNN.

INTRODUCTION

In this world every person is a social human being, who needs any mutual language to communicate with each other. But what about physically impaired people (both deaf and mute), whose language is not mutual to others. They communicate through their hand gestures and facial expressions which is not easily understood by others. To live in the society they require a converter (system) that converts their sign language to the common language. Sign Language Recognition (SLR) System fulfils need of the converter. It recognizes the hand gestures of deaf-mute people and converts it to the common language.

SLR System can be achieved in three simple steps :

Figure 1 SLR System

Data (hand gestures) is acquired using wearable sensors 2 and using vision based techniques. Vision based SLR are more convenient and easy to use. In vision based Data is acquired from images or videos. Data goes through preprocessing before sending to training. Hand regions are removed from the obtained image. Many of the researchers used different skin segmentation techniques to identify hand region in the image. RGB image is converted to different color spaces and then threshold value is applied to filter the skin pixels. But this technique limits to some specific skin tones. That type of SLR can be used in particular area where people of similar *skin tones live which is not good for its commercial use.*

In this paper, a general solution is proposed to overcome the problem of specific skin tone. It focuses on the method of acquiring data, irrespective of skin tone. Data is acquired using Region of interest 1 in which concept of background and foreground is used to identify hand region , and followed by connected component algorithm 9 to get hand shape more accurately. Then model is trained using Convolution Neural Network.

RELATED WORK

Sign Language Recognition has attracted the attention of many researchers as its area is wide. Much of the research work has been done in data acquisition part. Data is acquired by two techniques using wearable sensors 2 and by vision based techniques 9. Vision based techniques are more convenient as wearable sensors need extra care and are cost effective. In computer vision images obtained from camera are preprocessed before sending it to training. To identify hand region from whole image many researchers have used different skin segmentation techniques. 3 Used Kinect for depth and RGB images but it adds an extra hardware to the system which makes system more expensive. 6 Proposed SLR model that uses skin segmentation technique by converting RGB Image into YCbCr color space. 7 also converts RGB color space into YCbCr color space and followed by Convex Hull algorithm to determine the shape and location of hand. To ignore unwanted objects in image 4 proposed a Region of interest predictor and then applied skin segmentation technique by converting RGB Image into HSV color space . 5 used skin color segmentation and connected component algorithm to get hand region more accurately.

With the help of skin segmentation techniques in which RGB image is converted into different color space and then a threshold value is applied to filter skin pixels, one can easily identify skin region from the obtained image but it limits the model to predict the gestures of some specific skin tone that comes under threshold value.

To overcome the problem this paper proposes a method that identifies the hand region from obtained image ignoring its

skin color. With the help of Background Subtraction technique 11 is used which remove foreground object (hands) from the background that appears in Region of interest. It is followed by connected component algorithm to get hand shape more accurately.

PROPOSED MODEL

This work proposes a general solution to SLR of specific skin tones. The proposed model is divided into three parts:

1. Data Acquisition, 2. Training the Model, 3. Predict the Gesture. It focuses on foundation stone of the problem that is Data Acquisition. Since as accurately preprocessed data sent for training, the model would become more efficient in less data. In this work static gestures are taken into consideration .

Data Acquisition

Flow chart of data acquisition is given below:

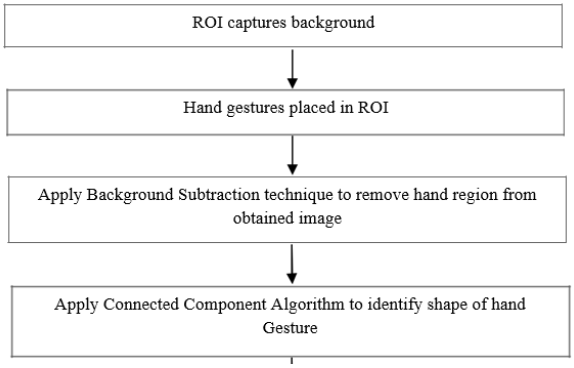

Figure 2 Flowchart of Data Acquisition

Data is acquired through ROI, to ignore unwanted objects 4 , from live video. Each frame of video acts as an image. Image requires preprocessing to maintain dataset of the model. ROI captures background for just few seconds and then static hand gesture is placed into the ROI which act as foreground object in image. Here concept of Background Subtraction technique 11 is used, which stated as

A pixel is defined as foreground pixel, if

$$[Bt - Bt+1] > t$$

where Bt is the pixel of frame taken at time t and Bt+1 is the pixel of frame taken at time t+1 from live video and t is predefined threshold value.

As result, hand region is identified and then followed by connected component algorithm to get hand shape more accurate. In Connected Component Algorithm 12 two pixels are marked as connected if they are of almost same pixel values (very less difference). After applying background Subtraction and connected component algorithm, shape of hand gesture is obtained . It results an image in which hand region is of white color and rest is of black color. Then this preprocessed image is sent for training the model.

Training the model

Model is trained by classifying the image. According to a review 8 SVM and HMM are frequently used techniques for image classification. But SVM give more accurate results in linear classification. Murat Taskiran et al [7 proposed a real time system that recognizes sign language using deep learning. This model is also trained using deep learning. Convolutional Neural Network is used in deep learning to classify the image into different classes. Architecture of CNN model used is given below:

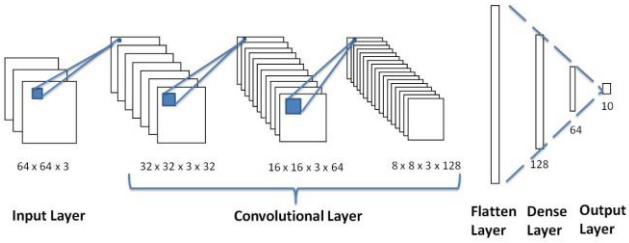

Figure 3 Architecture of CNN

It consists of 1 input layer, 3 convolutional layers, 1 flatten layer, 2 dense layers, 1 output layer Input layer consists of an RGB image. In Convolutional layer, 3×3 kernel is used to filter the important features from input image which is followed by max pooling of 2×2 to reduce dimensionality of input image. Flatten layer flattens the multi dimension data into single dimension and consists of 24,576 neurons. Dense layer also known as fully connected layer, each node is connected to the node in the next layer. 1st layer consists of 128 neurons and 2nd layer consists of 64 neurons. Output layer, it consists of 10 classes that means data is classified into 10 classes.

Predict the Gesture

Adam optimizer can be used to optimize the learning rate to improve the prediction rate.

CONCLUSION

This paper proposes the method of acquiring data with the combination of ROI, Background Subtraction and Connected Component Algorithm. In this way dependency of skin tone is removed and person of different skin tones can use this SLR.

FUTURE WORK

This paper presents the proposal hence its real time system is still required for its commercial use. A real time system can be developed by implementing the algorithms given in the proposal. Dataset should be made by acquiring data using given techniques in the paper. Model should be trained using

given architecture. Accuracy should be measured to compare the system than the proposed system 7.

REFERENCES

1. Sunmok Kim, Yangho Ji , and KI-Baek Lee, "An effective sign language learning with object detection based ROI segmentation" , IEEE Conference on Robotic Computing 2018.

2. Rinki Gupta, Arun Kumar, "Indian sign language recognition using wearable sensors and multi-label classification" ,Computers and Electrical Engineering, 2020.

3. Neel Kamal Bhagat, VishnuSai Y, Rathna G N, "Indian sign language gesture recognition using image processing and deep learning" , IEEE 2019.

4. SajanRaj T D, Beena M V, "Indian Sign Language Numeral Recognition using Region of Interest Convolutional Neural Network" , IEEE Conference on Inventive Communication and Computational

Technologies, 2018.

5. Sruthi C J, and Lijiya A, "Signet: A Deep Learning based Indian Sign Language Recognition System" , IEEE Conference on Communication and Signal Processing, 2019.

6. Resmi George, K Gerard Joe Nigel, "Hand Gesture Recognition for the application of Sign Language Interpretation" , International Journal of Engineering Research and Technology, 2014.

7. Murat Taskiran, Mehmet Killioglu, and Nihan Kahraman, "A Real-Time System for Recognition of American Sign Language by using Deep Learning", IEEE 2018.

8. Nimisha K P and Agnes Jacob, "A Brief Review of the Recent Trends in Sign Language Recognition", IEEE Conference on Communication and Signal Processing, 2020.

9. Thao Nguyen-Trang, "A New Efficient approach to detect Skin in Colour Image Using Bayesian Classifier and Connected Algorithm" , Mathematical Problems in Engineering, Hindawi 2018.

10. Kshitij Bantupalli, Ying Xie, "American Sign Language Recognition Using Deep Learning and Computer Vision" , IEEE Conference on Big Data, 2018.

11. Alan M. McIvor, "Background Subtraction Techniques".

12. Kurt Schwenk and Felix Huber, "Connected Component Labeling Algorithm for Very Complex and High Resolution Images on an FPGA Platform" , High Performance Computing in Remote Sensing, 2015.

13.