American Sign Language Recognition System

Balaji G
School of Computer Science and Engineering (SCOPE)
VIT CHENNAI
Chennai India
balaji.g@vitstudent.ac.in

1: 2nd Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 3rd Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

***Abstract*—In the field of multimodal communication, sign language is an area yet to be discovered completely. In line with the latest developments in the field of in-depth learning, there are far-reaching impacts and applications that can be made by neural networks in sign language recognition. In this paper, we introduce a way to use convolutional networks to identify images of both characters and digits in American Sign Language. Before going into the paper, sign language recognition has been done by many researchers, but our approach is quite different compared to others. Most of them used image processing techniques like using the HSV colour algorithm and making the background image black. Then they undergo many computer visions like grayscale, dilation and mask operation. Some of them used tools like Microsoft Kinect to process the images and extract the appearance-based feature and analyse the image in 2-D or 3-D format. Some of them used datasets to process the image, but they have used deep learning. In common all of these techniques consume a large amount of data and more processing power, which we students can't afford. As we are using the dataset, we don't have the necessity to train the images as there are numerous rows in the dataset which possess more efficiency in precision and accuracy. Also after using the CNN method, we are able to get better precision and accuracy.**

Keywords: Convolution neural network(CNN), American sign language(ASL), HSV(Hue Saturation Value), Microsoft Kinect.

I. INTRODUCTION

Sign language is generally considered as one of the unique forms of communication which are predominantly used for deaf peoples. The translating process between the normal spoken language and sign language is what is known as interpretation. Here, the interpretation function is the same as that of functions that we use for language translation.

Here, we use American sign language (ASL) which is predominantly used worldwide. There are 22 handshapes that correspond to the 26 letters of the alphabet, and on the other side, you can sign the 10 digits on one hand.

The manual method that we use in sign language is called fingerspelling. One of the reasons that the fingerspelling alphabet plays a vital role in sign language is that people used it to spell out names of anything for which there is not even a part of the sign. The main reason is that sign language is useful for people who are deaf and dumb. One of the examples where sign language is implemented is, in live news, there will be a separate screen where the newsreader uses sign language for the deaf and dumb people.

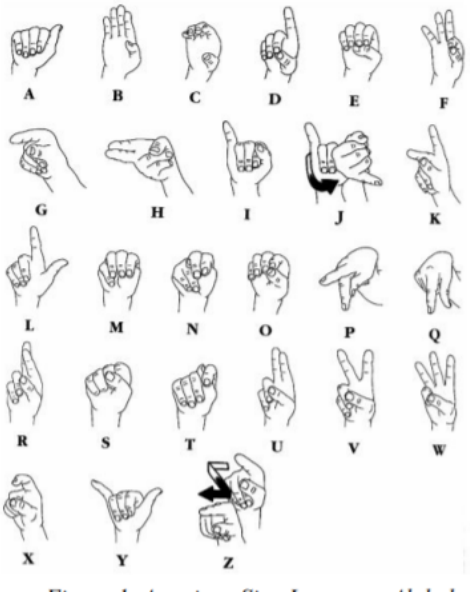


Fig.1: American sign language alphabet

II. MOTIVATION

Sign languages use visual cues to convey meaning. Sign languages are complete natural languages with their grammar and dictionary. Sign languages do not exist worldwide and do not understand each other equally, although there are significant similarities between sign languages.

Linguists consider both spoken and signed communications to be natural forms of language, meaning that both evolved into a vague ageing process, which took a long time and evolved without careful planning.

Wherever deaf communities are located, sign language has been developed as a useful means of communication, and they have become a major source of local deaf culture. Although signing is primarily used for the deaf and hard of

hearing, it is also used for hearing people, such as those who cannot speak physically, those who have language problems due to a disability or condition (additional communication etc.), or those with deaf family members, such as children of deaf adults.

So there are a large number of people who don't know sign language but live around these deaf people which creates a communication barrier. Also, even their needs cannot be satisfied since people cannot understand what they are trying to say. This drives us to relieve people out of this barrier.

LITERATURE REVIEW :

1)Paper details:

According to the World Health Organization (WHO) [1], the number of people with hearing or hearing impairments increased from 278 million in 2005 to 466 million in early 2018. [1]. This deaf community uses a collection of symbols to represent their language (called sign language), which differs from one nation to another. In other words, sign language (SL) is a non-verbal communication language, which uses patterns of visual gestures and any parts of the body, used mainly by people with hearing and/or hearing impairments. Sign languages (SLs) are fully natural language languages with their lexicon and grammar. Various SL versions have been developed, such as American Sign Language (ASL), Australian Sign Language, British Sign Language (BSL), Danish Sign Language, French Sign Language, and many more for deaf communities. Although there are significant similarities between the SLs, they are vague and uncommon. For example, ASL and BSL are different, although they both use the same voice language. Ordinary people who hear and listen find it difficult to understand sign language and even the nation itself. Therefore SL trained interpreters are required during medical and legal appointments, training and training sessions, etc.

Automatic SL recognition and its translation into the native language can establish appropriate communication between the deaf or hard of hearing and the general public. ASL is also a second language in deaf communities in the United States and Canada. According to the National Association of Deaf (NAD) in the United States of America, ASL is recognized by many high schools, colleges, and universities in fulfilling the requirements for modern and foreign qualifications. Outside of North America, ASL is used in many countries around the world, including parts of Southeast Asia and much of West Africa. There are other activities already reported in the books for automatic detection of ASL. Some of these methods have been analyzed in a sample database of a few samples, while others use a neural network method of shallow differentiation. Deep neural networks require manual manipulation of features and selection of appropriate features.

The use of deep learning (DL) techniques for machine learning problems has greatly improved the performance of traditional neural networks, especially image recognition and computer vision problems. DL is a subset of machine learning in artificial intelligence (AI). It is a collection of algorithms, models with a high degree of abstraction with structures built with many indirect mutations. DL algorithms use large amounts of data to generate features automatically, aiming to mimic the human brain's ability to read, analyze, view, and visualize, especially in the most complex problems. DL structures build relationships beyond the immediate neighbours of the data and produce learning patterns, delivering direct presentations to the data without human intervention. There are different types of deep learning structures such as deep belief networks, embedded automatic coding, convolutional neural networks, and so on.

Among them, CNN has used the multi-layer artificial neural network (ANN) to provide modern precision in the field of computer vision, medical image analysis, speech recognition, bioinformatics, and more. Convolutional neural network (CNN), one of the most well-known in-depth study

PROPOSED WORK

In this proposed work, our goal is to recognize American sign language using various image processing techniques and adding image processing techniques with CNN(Convolutional neural network) to make a real-time system.

CASE 1 USING MMATLAB

RGB colour

The RGB colour model is an additional colour model where the main light colours red, green, and blue are added together in a variety of ways to reproduce a wide range of colours. The name of this model comes from the three main colour letters to add, red, green, and blue.

The main purpose of the RGB colour model is to hear, represent, and display images on electronic devices, such as televisions and computers, although it has also been used for conventional photography. Before the electronic era, the RGB colour model already had a strong vision behind it, based on human perception of colours.

RGB device-dependent colour model: different devices receive or reproduce the RGB value given differently, as colour factors (such as phosphors or dyes) and their reaction to the same red, green, and blue levels vary from manufacturer to manufacturer, or on the same device over time. So RGB value does not mean the same colour on all devices except for some type of colour management.

The YCbCr Color Space

The YCbCr colour space is widely used in digita lvideo, image processing, etc. In this format, light details are represented by one part, Y, and colour information are kept in pairs parts of colour differences, Cb and Cr. Element Cb is the difference between the blue part and reference number, and part Cr difference between the red part and the reference number.

The transformation used to convert from RGB to YCbCr colour space is shown in equation (1):

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112 \\ 112 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

In contrast to RGB, the YCbCr colour space is lumaindependent, resulting in a better performance. The corresponding skin cluster is given as:

$$Y > 80$$
$$85 < Cb < 135$$
$$135 < Cr < 180,$$
$$\text{Where } Y, Cb, Cr = [0, 255].$$

Chai and Ngan developed that algorithm takes advantage of the local features of human skin colour. Skin colour map is taken and applied to chrominance components of the input image to be obtained the pixels look like skin. Working in this colour space Chai and Ngan found that range Cb and Cr are many skin colour representatives a reference map is available:

77≤Cb≤127 and 133≤Cr≤173

However, our aim is to determine the skin of a person of different races, the boundaries are given above apply only to Caucasian skin because the first border only treats people with white skin, and the second limit separates people from different parts of the world but some pixels are found as skin but not. For this reason, a new skin boundary is proposed to separate the people within the image.

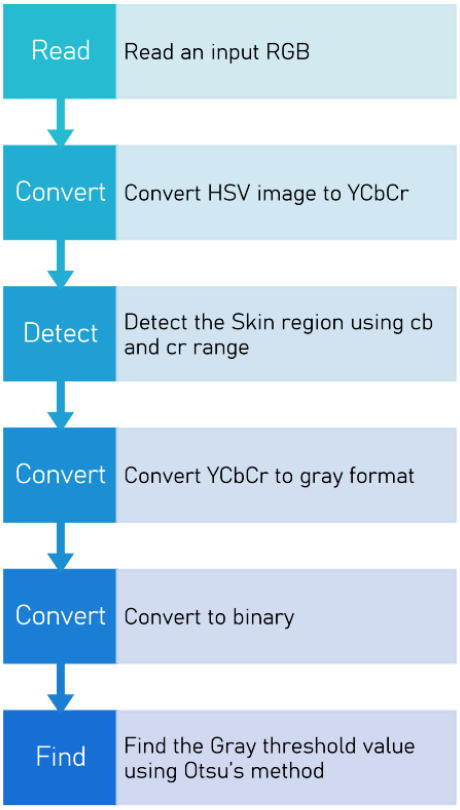
80≤Cb≤120 and 133≤Cr≤173

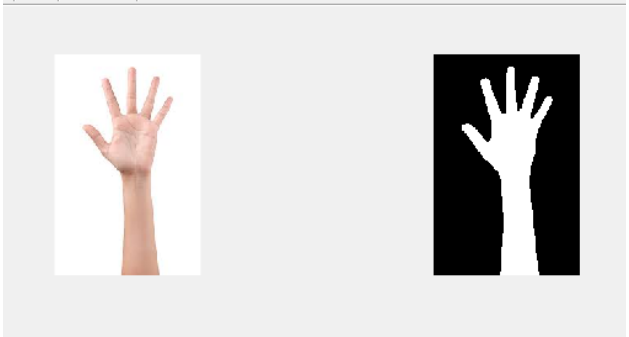
Skin Detection

Skin discovery can help in finding a human organ, body, or face inside the image. Recently many skin types identification within a digital image has been improved. Skin colour proved to be useful to a solid way to get a face, local and once to track. There have been several researchers who have been looking to use colour information to find out skin.

Grey threshold value using Otsu's method

The algorithm returns one limit that divides pixels into two classes, front and back. This limit is determined by minimizing the variability in class intensity, or equally, by increasing class variability. Otsu's method is a clear one-sided analogy of Fisher's Discriminant Analysis, related to the Jenks method, and equals the world's best k-methods performed in a solid histogram. Multi-level expansion is described in the original paper, and a computer-based implementation has since been proposed.





Step 1: Start

Step 2: Obtain the Centroid (Central mass of the region) of the preprocessed image.

Step 3: Find the Area (actual number of pixels in the region) and Perimeter (the distance around the boundary of the region) of an image.

Step 4: Equation (1) is used to find the Roundness and the Boundary of the preprocessed image.

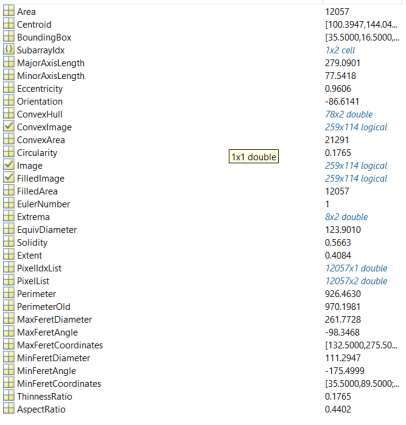
$$RND = \frac{(4 * \pi * ARE)}{PER^2} \quad (1)$$

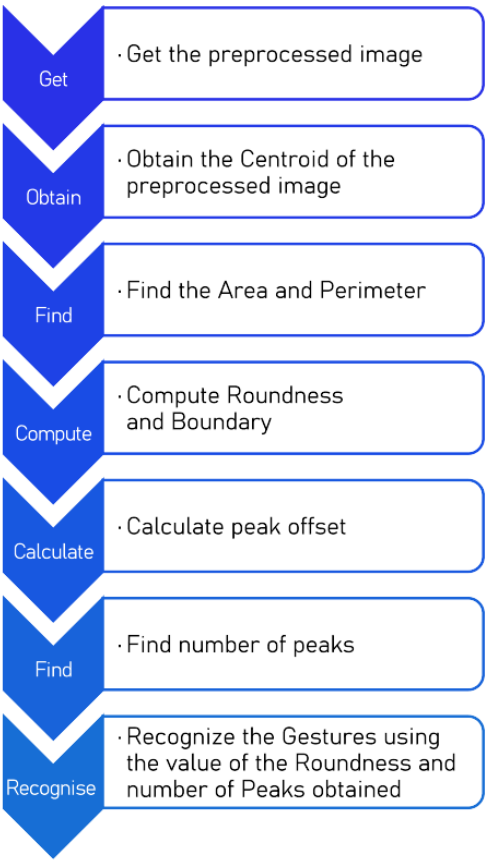
Step 5: Equation (2) is used to compute the Peak Offset point from the Centroid.

$$pkoffset = CEN(:,2) + 0.9 * (CEN(:,2)); \quad (2)$$

Step 6: Find the number of Peaks.

Step 7: Recognize the Gestures using the value of the Roundness and number of Peaks obtained.





CASE 1 USING MMATLAB

Image Acquisition Model:

Image acquisition is the process of creating photographic images, such as the interior structure of an object. The term is often assumed to include the compression, storage, printing, and display of such images. This is a primary and essential step in sign recognition. Camera interfacing is a necessary task to capture images with the help of Webcam. Nowadays lots of Laptops are coming with inbuilt camera systems so that helps a lot for capturing images to process it further. Gestures can be captured by an inbuilt camera to detect hand movements and position. Capturing 30fps will be sufficient to process images; more input images may lead to higher computational time and will make the system slow and vulnerable.

DATA SET:

Here we are using images to create the data set. Creating a directory and sub directory with the name of character or number of the sign. Which will be further used to create the training dataset used to create CNN model.

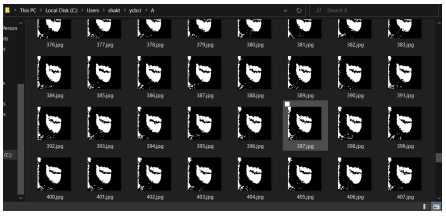


Image Processing model:

While doing research we used four techniques to find which is best for the real-time machine. We are using the below colour spaces and filters :

- YCrCb Color Space
- HSV Color Space
- Edge detection using filters
- HSV with Histogram back projection

Upon trying with the above 4 YCbCr colour model gave us the best results of 90% accuracy, whereas HSV 86%, edge detection 71% and HSV with histogram back-projection 81%.

```
[16]: score = model.evaluate(test_x, test_y)
      print('Test accuracy: %0.2f%%' % (score[0] * 100))
      print(color_model)

      Test accuracy: 90.00%
      1
```

YCrCb Color Space:

The YCbCr colour space is widely used in digital video, image processing, etc. In this format, light details are represented by one part, Y, and colour information are kept in pairs parts of colour differences, Cb and Cr. Element Cb is the difference between the blue part and reference number, and part Cr difference between the red part and the reference number.

The transformation used to convert from RGB to YCbCr colour space is shown in equation (1):

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112 \\ 112 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

In contrast to RGB, the YCbCr colour space is lumaindependent, resulting in a better performance. The corresponding skin cluster is given as:

$$Y > 80$$
$$85 < Cb < 135$$
$$135 < Cr < 180,$$
$$\text{Where } Y, Cb, Cr = [0, 255].$$

Chai and Ngan developed that algorithm to take advantage of the local features of human skin colour. Skin colour map is taken and applied to chrominance components of the input image to be obtained the pixels look like skin. Working in this colour space Chai and Ngan found that range Cb and Cr are many skin colour representatives a reference map is available:

77≤Cb≤127 and 133≤Cr≤173

However, we aim to determine the skin of a person of different races, the boundaries are given above apply only to Caucasian skin because the first border only treats people with white skin, and the second limit separates people from different parts of the world but some pixels are found as skin but not. For this reason, a new skin boundary is proposed to separate the people within the image.

80≤Cb≤120 and 133≤Cr≤173

HSV Color Space:

The HSV colour space is more sensitive to the way people feel colour than the RGB colour space. As colour (H) varies from 0 to 1.0, the corresponding colours vary from red to yellow, green, blue, blue, and magenta, after red. Since filling (S) varies from 0 to 1.0, matching colours (hues) vary unsaturated (grey shades) to fill (no white part). As a value (V), or light, it varies from 0 to 1.0, I the corresponding colours are getting brighter

The first step in skin tanning is to convert the captured image to HSV colour Space is the most suitable for our study. The HSV colour model is a cylinder representation of a typical RGB model. HSV stands for Hue, Saturation, and Value. Hue is measured in degrees and varies from 0 to 360. Create a basic colour. The filling of the space and the amount (light) determines the approach of white and black respectively. In the basic model, they vary from 0 up to 100, but in the OpenCV library used on the detector, they vary from 0 to 255. To convert an image from RGB to HSV, each pixel in the image is subject to the following version

The HSV value for each pixel is compared to standard skin pixel values and a decision is made whether the pixel is a skin pixel or not depending on the values in the range of the predefined range values for each parameter.

The pixel width of different colored areas used by our algorithm is as follows:

0.0 <= H <= 50.0 and 0.23 <= S <= 0.68

Edge detection using filters:

Laplacian Operator is also a derivative operator which is used to find edges in an image. The major difference between Laplacian and other operators like Prewitt, Sobel, Robinson and Kirsch is that these all are first-order derivative masks but Laplacian is a second-order derivative mask. In this mask we have two further classifications one is Positive Laplacian Operator and the other is Negative Laplacian Operator.

Another difference between Laplacian and other operators is that unlike other operators Laplacian didn't take out edges in any particular direction but it take out edges in the following classification.

HSV with Histogram back projectioThe originalinal algorithm was proposed by Swain and Ballard in their article "Color Indexing" in 1991. Backprojection answer to the question "Where in the image are colourslors that belong to the object being looked for (the target)?" Given the image histogram I and the model histogram M, we define a ratio histogram R as: $R_i = min(\frac{M_i}{I_i},1)$ function h(c) maps color c of the pixel at (x,y) to the value of the

histogram R it indices. In other words, function hback projectsects the histogram R onto the input image. back-projectedcted image B is then convolved with a disk D of radius r.

Preprocessing ModeThe main

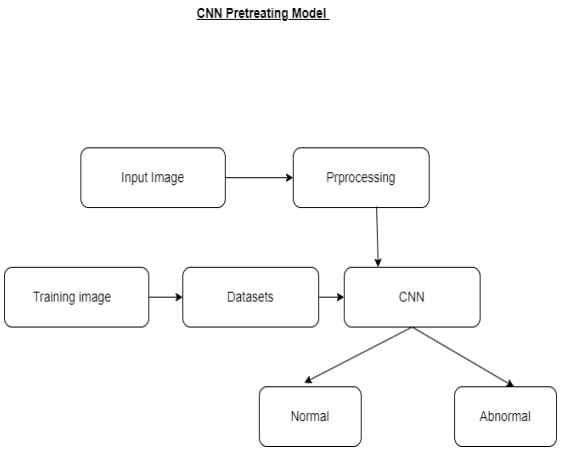
Main aim of pre-processing is an improvement of the image data that reduces unwanted deviation or enhances image features for further processing. Preprocessing is also referred to as an attempt to capture the important pattern which expresses the uniqueness in data without noise or unwanted data which includes cropping, resizing greygray scaling. Cropping refers to the removal of the unwanted parts of an image to improve framing, accentuate subject matter or challenge aspect ratio. Resizing Images are resized to suit the space allocated or available. Resizing images are tips for keeping the quality of the original image. Changing the physical size affects the physical size but not the resolution.

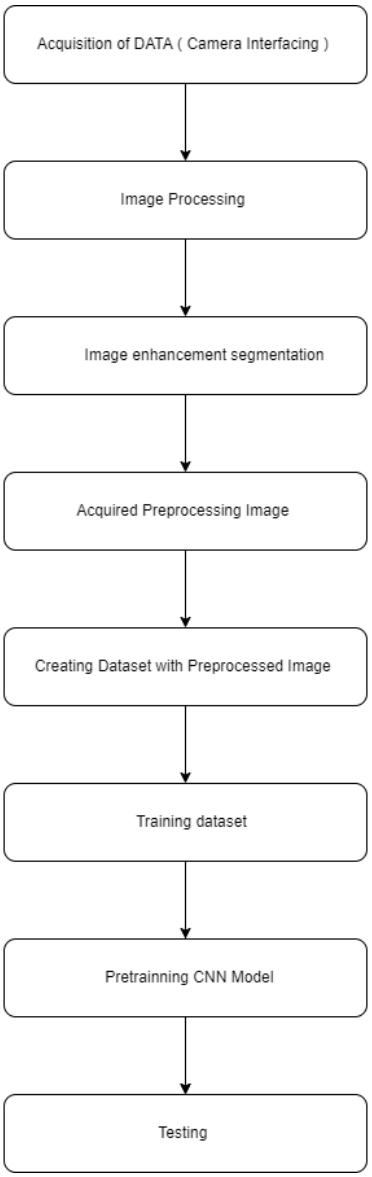
neural networks CNNs are trained with a version of the backpropagation algorithm. Convolutional layer :Core building block of a CNN. Layer's parameters consist of a set of learnable filters (or kernels). Filter is convolved across the width and height of the input volume. Computing the dot product between the entries of the filter Pooling Layer: Reduce the spatial size of the representation to reduce the amount of parameters. Independently operates on every depth slice of the input. Most common form is a pooling layer with filters of size 2x2 applied with a stride of 2 down samples every depth slice in the input by 2 along with both the width and the height, discarding 75% of the activations spatially, using the MAX operation

ReLU layer: ReLU is the abbreviation of Rectified Linear Units which increases the nonlinear properties

Fully connected layer:Neurons in a fully connected layer have full connections to all activations in the previous layer. The activations are computed with the matrix multiplication.

We are using 7 layered CNN where the final layer is a regression equation since the hand signs are identified with numbers1-28 including blank1.





Conclusion:

This proposal will help to achieve high performance in recognizing the sign language, which is the main communication bridge between the deaf and dumb people and the normal people. It is hard for most of the people who are not familiar with sign language to communicate without an interpreter. In this proposal, we have created an idea of translating the static image of sign language to the spoken language of hearing. The static image includes the alphabet and some words, used in both training and testing of data. Feature representation will be learned by a technique known as convolutional neural networks. The new representation is expected to capture various image features and complex non-linear feature interactions. A softmax layer will be used to recognize signs.

Abbreviations and Acronyms

CNN-convolutional neural network(A convolutional neural network (CNN) is a type of artificial neural network used in image recognition and processing that is specifically designed to process pixel data. CNNs are powerful image

processing, artificial intelligence (AI) that use deep learning to perform both generative and descriptive tasks, often using machine vision that includes image and video recognition, along with recommender systems and natural language processing (NLP).)

Acknowledgement

I would like to express my special thanks of gratitude to my teacher (Dr Geetha S) who gave me the golden opportunity to do this wonderful project on the topic (American Sign Language Recognition System), which also helped me in doing a lot of research and I came to know about so many new things I am really thankful to them.

Secondly, I would also like to thank friends and teammates who helped me a lot in finalizing this project within the limited time frame.

References

[1]Srinath S, Ganesh Krishna Sharma, “Classification approach for Sign Language Recognition”, International Conference on Signal, Image Processing, Communication & Automation, 2017.

[2]Srinath S, Ganesh Krishna Sharma, “Classification approach for Sign Language Recognition”, International Conference on Signal, Image Processing, Communication & Automation, 2017

[3Sandor, DielemanPieter-Jan Kindermans, Benjamin Schrauwen, “Sign Language Recognition Using Convolutional Neural Networks”

[4]Jayshree R. Pansare, Maya Ingle, “Vision-Based Approach for American Sign Language Recognition Using Edge Orientation Histogram”, International Conference on Image, Vision and Computing, pp.86-90, 2016.

[6]Mehreen Hurroo, “Sign Language Recognition System using Convolutional Neural Network and Computer Vision”