INDIAN INSTITUTE OF INFORMATION TECHNOLOGY, ALLAHABAD

INFORMATION TECHNOLOGY DEPARTMENT

SIGN LANGUAGE DETECTION

Under the guidance of Dr.Triloki Pant

GROUP MEMBERS:

Jaya Meena(IIT2018029), Ankita Chandra(IIT2018053),
Manisha Kumari(IIT2018062), Aastha Kumari(IIT2018091),
Trupti Pendharkar(IIT2018097)

Abstract- Sign languages (also known as signed languages) are languages that use the visual-manual modality to convey meaning. Sign language is one of the oldest and most natural forms of language for communication, but since most people do not know sign language and interpreters are very difficult to find in day to day conversations, we have come up with a real time method for fingerspelling based Indian sign language. In our method, the hand is first passed through a median blur filter and canny edge detection algorithm is applied to the filtered image, then feature extraction using SURF is performed on the result and further model of visual words are obtained after clustering, and then svm classifier is trained on histogram computed through these visual words to generate our model. The method used provides 99 % accuracy for the 26 letters of the alphabet and 1 to 9 numbers.

I. PROBLEM DEFINITION

Sign Language Detection project is based on the real life problems for deaf and dumb people who use sign language to communicate. There are very few people who can understand sign language and thus makes it difficult for the deaf and dumb people to communicate. In this project, we are going to recognize sign language using hand gestures which will make it easier for the handicapped people to communicate with people who do not understand sign language. The aim is to build a human computer interface which can solve the above described problem in the simplest way possible, and with great accuracy.

II. INTRODUCTION

Sign Language is the oldest and the natural form of language for communication. The process of exchange of thoughts and messages in various ways such as speech, signals, behaviour and visuals is called communication. Deaf and dumb people use their hands to express their ideas. Gestures are non verbally exchanged messages and are understood with vision. A sign language is like any other language which has vocabulary and grammar.

Indian sign language is a predominant sign language since the only disability D&M people have is communication and they cannot use verbal languages hence the only way for them to communicate is through sign language. Communication is the process of exchange of thoughts and messages in various ways such as speech, signals, behavior and visuals. Deaf and dumb(D&M) people make use of their hands to express different gestures to express their ideas with other people. Gestures are the nonverbally exchanged messages and these gestures are understood with vision. This nonverbal communication of deaf and dumb people is called sign language. In our project we basically focus on producing a model which can recognise Fingerspelling based hand gestures in order to form a complete word by combining each gesture. The gestures we aim to train are as given in the image below.

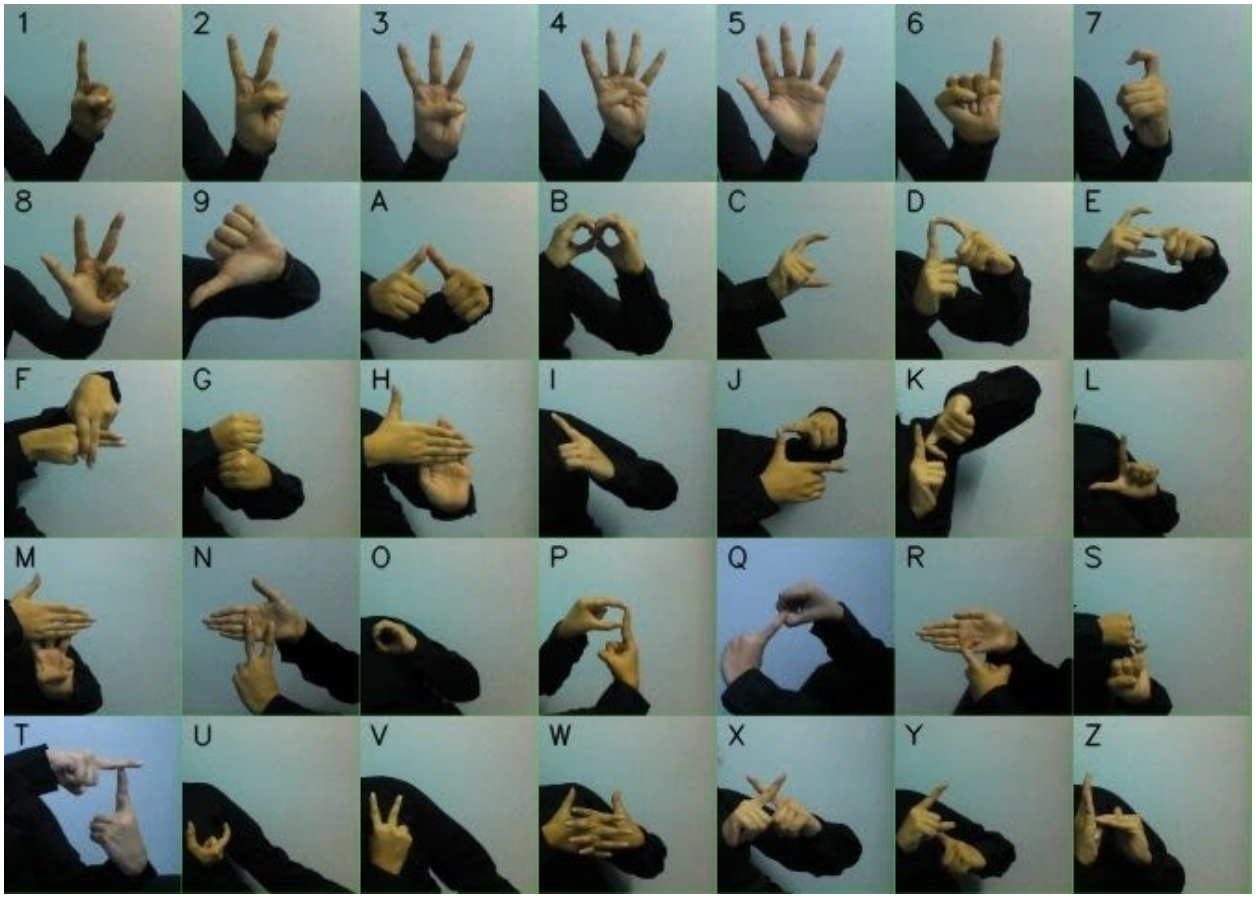


Fig 1: ISL gesture for 26 alphabets and numericals from 1-9.

III. MOTIVATION

Communication is a basic requirement for survival in society. The sign language is a form of communication using hands,limb,head as well as facial expression which is used in a visual and spatial manner to communicate without sound. It play a major role for Deaf and dumb people who communicate among themselves using sign language but normal people find it difficult to understand their language. Sign language poses the challenge that they are multi-channel; conveying meaning through many modes at once. Sign language involves many features which are based around the hands, in general there are hands shape/orientation and movement trajectories, which are similar in principle to gestures. While many gestures recognition techniques are applicable , Sign language provides a more complex challenge than the traditionally more confined domain of gesture recognition. Our project aims at taking the basic step in reducing the communication gap between normal people and deaf and dumb people using Indian sign language. Effective augmentation of this project to words and common expressions may not only help the deaf and dumb people communicate faster and simpler with outer world, but also encourage in developing autonomous systems for understanding and helping them.

IV. LITERATURE SURVEY

Little work has been done before on ISL. Sakshi Goyal, Ishita Sharma implemented feature detection using SIFT technique and then matching the keypoint of a new image with the key points of trained images per alphabet in a database to classify the new image with the label of one with the nearest match.
Subhash Chand Agrawal et al implemented the recognition of two-handed Indian sign language (ISL). Their method implementation contains 3 steps namely Image segmentation, Feature extraction, Recognition. They used the Otsu algorithm for Image Segmentation. For key feature extraction they used SIFT technique and HOG descriptors. For Classification, they used a multi-class Support Vector Machine(MSVM) model and trained it with all features.. Our method is also similar to them, but we used the canny edge algorithm and SURF algorithm for feature extraction instead of SIFT in the Image segmentation phase.

Data Acquisition

There are mainly two approaches to acquire data about the hand gesture:

- **Sensory devices**
  We can use sensory devices to provide us data about the exact configuration and position of the hand by using electromechanical devices (glove based approaches). But these devices are expensive and not user friendly.

- **Vision based devices**
  We can use different vision based input devices (camera) for observing the configuration and position of the hands and the fingers. These systems complement biological versions due to an artificial system that is implemented in the software, hardware systems. This approach is user friendly since it only requires a camera and it is also quite user friendly . The challenge for this system is to handle a large variety of hand's appearance due to hand movement and different skin tones and also variation in view points , scales and also the speed of the camera capturing the gestures or any scene.

Due to the lack of resources like electromechanical gloves and their expensive nature we will stick to the vision based method that only needs a vision based input device, i.e. camera/ web camera.

Feature Extraction

- SURF stands for Speeded Up Robust Feature is a patented local feature detector and descriptor.

SURF is a novel feature extraction method which is robust against rotation, scaling, occlusion and variation in viewpoint.

The SURF algorithm has three steps:

1. **Interest point detection:**
   SURF uses an integer approximation of the determinant of Hessian blob detector, which can be computed with 3 integer operations using a precomputed integral image.
2. **Local neighborhood description:**
   Check the intensity distribution of the pixels within the neighbourhood of the point of interest
3. **Matching:**
   By comparing the descriptors obtained from different images, matching pairs can be found.

- We also calculate histograms using the predicted visual words. Histogram can be calculated by finding the frequency of occurrence of each visual word that belongs to an image in total visual words.

And also the amount of data to train is huge and many gestures look similar and this technique will not provide good accurate results. To avoid it instead of segmenting the hand out of the different background we will make every background of the hand a stable single color so that segmentation on the basis of skin can be avoided

Gesture Classification

- **Hue, Saturation, Value (HSV)**

  This model deals with the dynamics of the gesture. HSV is a cylindrical color model that remaps the RGB primary colors into dimensions that are easier for humans to understand.

  The gestures are extracted by tracking the skin-color blobs corresponding to the hand into a body and adding face space centered on the face of the user. This result should belong to any one of the classes of the gestures, i.e deictic class and symbolic class. A fast look-up indexing table is used for filtering the images and after that all skin color pixels are collected in blobs.

- **K-Means Clustering**

  In this algorithm the idea is to define k centers for each cluster. Then take the next point in the dataset and associate it with the nearest neighboring cluster's center. This process is repeated until

  The centroids have stabilized, i.e. there is no change in their values because the clustering has been successful or the defined number of iterations has been achieved.

SVMs are a set of supervised learning methods used for classification, regression and outliers detection.

There are several advantages associated with SVMs

1. They are effective in High dimensional space.
2. Even if no. of dimensions are greater than total samples they are effective.
3. They are memory efficient.
4. Main advantage is their versatility as different kernel functions can be used for decision functions.

V. KEY WORDS AND DEFINITIONS

Feature extraction and representation: An image is represented in 3D matrix with dimensions as width and height and depth which is the value of each pixel. The depth is 1 in case of grayscale image and 3 in case of RGB image. These pixel values are used for extracting features using Bow model and K means clustering .

SURF: SURF stands for Speeded Up Robust Feature is a patented local feature detector and descriptor.

VI. METHODOLOGY

Project is based on the concept of computer vision and it also eliminates the problem of using any artificial device for interaction as all the signs or gestures are represented with bare hands.

DatasetGeneration

As less research has been done for the Indian Sign Language as compared to ASL proper dataset is not available for ISL, so we have prepared our own dataset. We have built a python file through which we can generate our data for all the Classes. So for creating a dataset we have to use the Open Computer Vision(OpenCV) library. Firstly we captured around 7000 total images 200 for each 35 labels ISL. Then we divided the dataset in 80:20 percent ratio into training and testing data respectively.

Gesture Classification

Our Approach for Sign language classification:
Our approach used following steps for final detection of symbol

Algorithm Step 1: Image Segmentation

Algorithm Step 2: Feature extraction

Algorithm Step 3: SVM Model for Classification

Step 1: Image Preprocessing

Image Segmentation

The goal of Image segmentation is to remove background and noises or we can say simplify and/or change the representation of an image into something which is Region of Interest (ROI) and the only useful information in the image. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images.

Image segmentation can be achieved when following steps are performed:

1. Skin Masking: Using the concept of thresholding this RGB color space is converted into grayscale image and SkinMask is finally obtained through HSV color space(which we get from gray scale image)

2. Canny edge detection: It is basically a technique which identifies or detects the presence of sharp discontinuities in an image there by detecting the edges of the figure in focus.

It is multi step algorithm which is followed as:

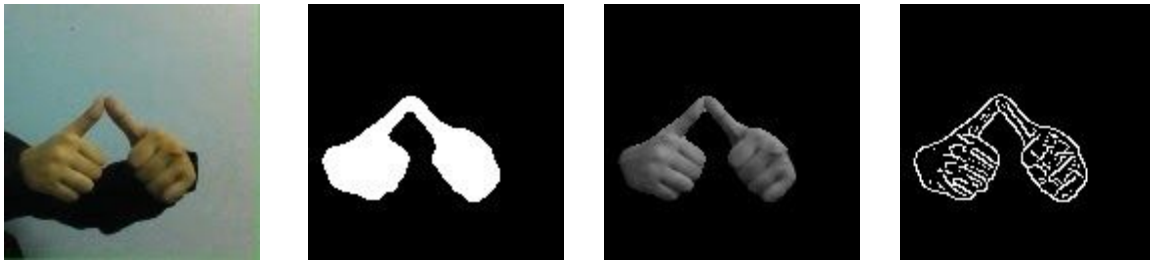
Step 1: Computing the horizontal (Gx) and vertical (Gy) gradient of each pixel in an image.

Step 2: Using the above information the magnitude (G) and direction (of each pixel in the image is calculated.

Step 3: In this step all non-maxima‟s are made as zero that is suppressing the non- maxima‟s thus the step is called Non-Maximal Suppression.

Step 4: The high and low thresholds are measured using the histogram of the gradient magnitude of the image

Step 5: To get the proper edge map hysteresis thresholding is employed which will link between the weak and strong edges. The weak edges are taken into consideration if and only if it is connected to one of the strong edges or else it is eliminated from the edge map. The strong edge is the one whose pixel is greater than the high threshold and weak edge is one whose pixel value lies between high and low threshold.



Raw Image | Skin Mask(gray scale) | Skin | Canny Edge detection

Fig 2: Different stages of Image Preprocessing.

Step2: Feature Extraction

In feature extraction following steps were performed:

1. Feature detection : key features of the image were extracted using SURF technique. SURF is a feature extraction algorithm which is robust against rotation variation scaling. We have extracted features using the inbuilt SURF function in opencv.

2. Clustering: To cluster all the features obtained in the above step we apply mini batch k-means clustering(similar to K-means clustering but efficient in terms of time consumption and memory).

   In our code we have taken k as 8 for the image of each label. So the cluster size or total number of feature descriptors are 8 * 35. After training of all SURF features (extracted in above step) through mini batch k-means clustering, all similar features are clustered in a cluster. Total number of clusters are also known as visual words.
   So in this step we obtained visual words for each image.

3. Histogram Computation: In this step we computed Histogram using predicted visual words(generated above). This is done by calculating the frequency of each visual word belonging to the image in total visual words.

Step 3: Classification

SVM Model for Classification:

Once all the histograms are generated for the total data set using the above step, the training dataset is trained using **Support Vector Machine Classifier** and then predicted with a linear kernel. Other Classifiers like CNN, KNN, Logistic Regression can also be used for classification.

Additionally for the purpose of real time recognition the trained model is saved in a file so that a user can predict the gesture using video feed in real time.

VII. TRAINING AND TESTING

First we convert our data from RGB to grayscale then after applying medianblur filter(to get skin masked image) canny edge detection algorithm is applied to all images using inbuilt Canny method. After this SURF algorithm is applied to all images to get feature descriptors.

Once feature descriptors are obtained for all images these are trained over the clustering model, and histograms are computed, these training data histograms mapped with their respective labels and are trained over SVM classifier and once training is done histograms generated labels for testing dataset are predicted using SVM classifier predict method.

VIII. CHALLENGES FACED

There were many challenges faced by us during the project. The very first issue we faced was the dataset. There were no stable datasets on the web so we had to create our own dataset. And we captured images wearing full sleeves dark color clothes in order to reduce noise or extract necessary information from the image. And we have to select a specific background for the same.

IX. RESULTS

We scored a 99% accuracy using the Bow, integrated with robust SURF feature descriptors. The real time recognition prediction of results can be seen by the figure below. Using a large set of data images always helps us to get a better efficiency in result as there could be slight biasing in the model prediction as the data set has much similar images without variations.

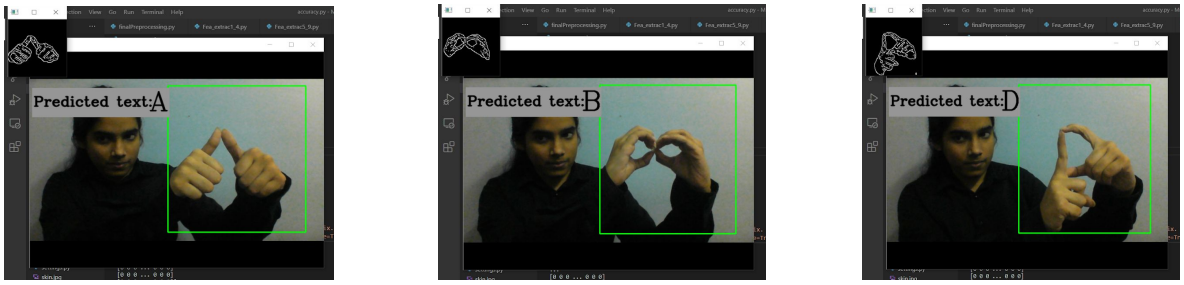


Fig 3: Real time Prediction for the **A**,**B** and **D** alphabet is shown in the above images.

We also created a file to generate the confusion matrix for our model which is shown in below figure

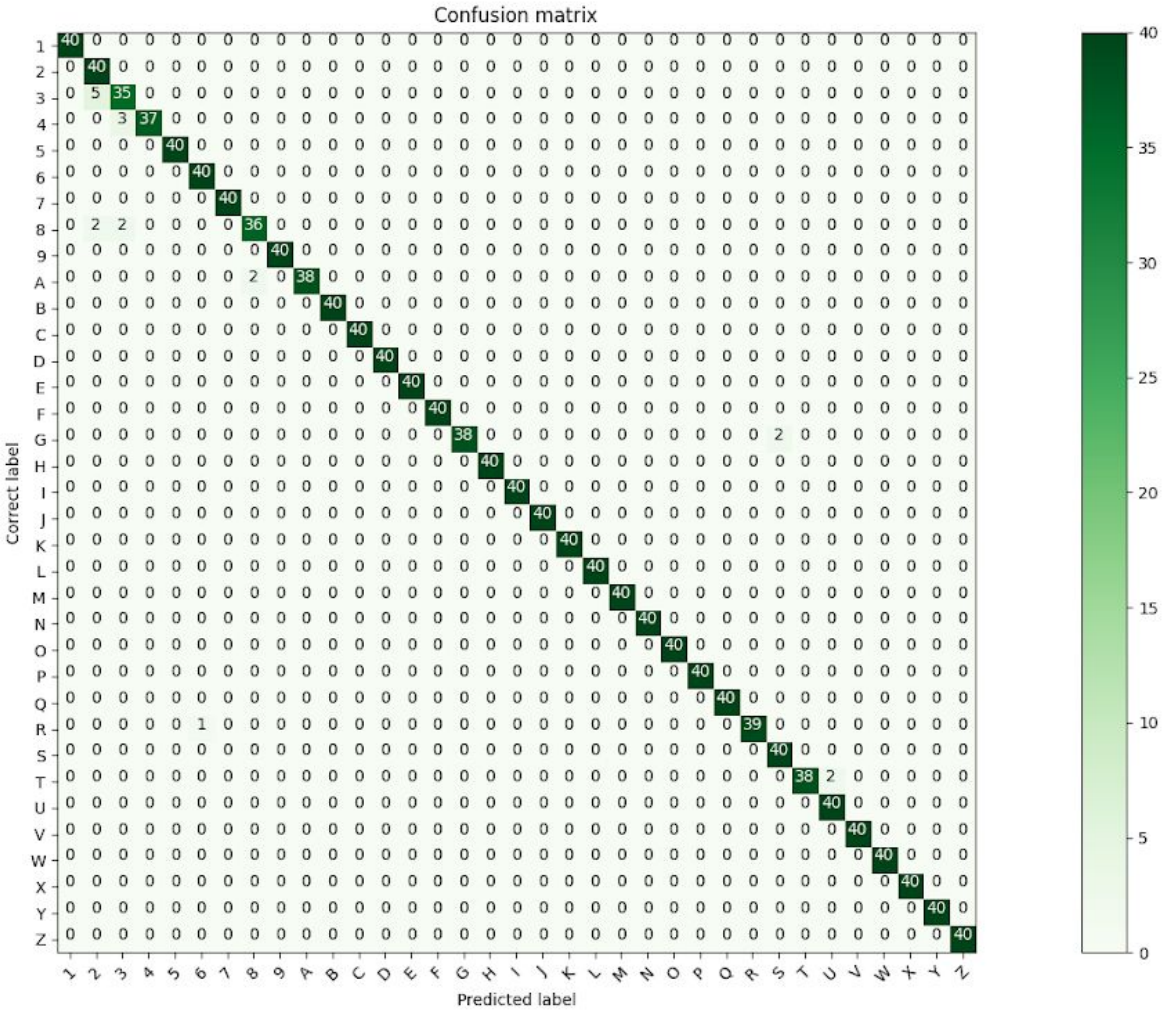


Fig 4: Confusion matrix for our model.

We can see that almost 99 % of labels are predicted correctly. Some labels with incorrect prediction are G,T,3,4 which are wrongly predicted with S,U ,2,3respectively.

The 40 in the confusion matrix represents the total number of testing dataset for each label which is 20% of 200.

Metrics showing Precision score, Recall score and F1 score for each label are attached as follows

label	True Positive	False Positive	Precision	False negative	Recall	F1 Score
1	40	0	1	0	1	1
2	40	7	0.85106	0	1	0.91954
3	35	5	0.875	5	0.875	0.875
4	37	0	1	3	0.925	0.96104
5	40	0	1	0	1	1
6	40	1	0.97561	0	1	0.98765
7	40	0	1	0	1	1
8	36	2	0.94737	4	0.9	0.92308
9	40	0	1	0	1	1
A	38	0	1	2	0.95	0.97436
B	40	0	1	0	1	1
C	40	0	1	0	1	1
D	40	0	1	0	1	1
E	40	0	1	0	1	1
F	40	0	1	0	1	1
G	38	0	1	2	0.95	0.97436
H	40	0	1	0	1	1
I	40	0	1	0	1	1
J	40	0	1	0	1	1
K	40	0	1	0	1	1
L	40	0	1	0	1	1
M	40	0	1	0	1	1
N	40	0	1	0	1	1
O	40	0	1	0	1	1
P	40	0	1	0	1	1
Q	40	0	1	0	1	1
R	39	0	1	1	0.975	0.98734
S	40	2	0.95238	0	1	0.97561
T	38	0	1	2	0.95	0.97436
U	40	2	0.95238	0	1	0.97561
V	40	0	1	0	1	1
W	40	0	1	0	1	1
X	40	0	1	0	1	1
Y	40	0	1	0	1	1
Z	40	0	1	0	1	1

Fig: 5 Accuracy metrics

X. CONCLUSION

Indian sign language is a principal for communication for deaf and dumb people in India. This paper gives a complete implementation for Indian sign language recognition using the Bag of words model. In section 6, step wise implementation has been discussed which are image collection, image pre-processing, feature extraction (using K-means clustering, visual words collection) and Classification. There is not only development of static images by recognition , but there is also a development of a real time recognition of gestures . This project can also be extended for simple expressions and words in ISL including alphabets and numeric.

XI. FUTURE SCOPE

We are planning to achieve higher accuracy even in case of complex backgrounds by trying out various background subtraction algorithms . The Image Processing part should be improved so that the System would be able to communicate in both directions i.e. it should be capable of converting normal language to sign language and vice versa. Moreover we will focus on converting the sequence of gestures into text i.e. word and sentences and then converting it into speech which can be heard.

XII. REFERENCES

[1] Shravani K,et al. "Indian Sign Language Character Recognition." IOSR Journal of Computer Engineering (IOSR-JCE), 22(3), (2020), pp. 14-19.

[2] S. C. Agrawal, A. S. Jalal, and C. Bhatnagar. Recognition of indian sign language using feature fusion. In 2012 4th International Conference on Intelligent Human Computer Interaction (IHCI), pages 1–5, 2012.

[3] Joyeeta Singh, K. D. Indian sign language recognition using eigen value weighted euclidean distance based classification technique. International Journal of Advanced Computer Science and Applications 4, 2 (2013).

[4] Neha V. Tavari, P. A. V. D. Indian sign language recognition based on histograms of oriented gradient. International Journal of Computer Science and Information Technologies 5, 3 (2014), 3657–3660.

[5] https://research.ijcaonline.org/volume70/number19/pxc3887306.pdf

[6] Sakshi Goyal, Ishita Sharma, S. S. Sign language recognition system for deaf and dumb people. International Journal of Engineering Research Technology 2, 4 (April 2013).