EasyChair Preprint

№ 8133

Sign Language Recognition System

Akriti Goyal, Deepanshu Dhar, Paras Nair, Chirag Saini,
Chetana Prakash and S M Supreetha

May 31, 2022

Sign Language Recognition System

Akriti Goyal, Deepanshu Dhar, Paras A Nair, Chirag Saini, Dr. Chetana Prakash, Prof. Supreetha S M

Department of computer science and engineering, BIET Davangere

Abstract: ***Communication plays one of the important role in survival within the society. People with ability to speak and hear are privileged one. Dumb and deaf people are usually deprived of normal communication with other people and find it really difficult at times to interact with normal people with their gestures, as only a very few of those are recognized by most people. Sign Language is the primary means of communication in the deaf and dumb community. As like any other language it has also got grammar and vocabulary but uses visual modality for exchanging information. The problem arises when dumb or deaf people try to express themselves to other people with the help of these sign language grammars. As a result, it has been seen that communication of a dumb person is only limited within his/her family or the deaf community.***

Keywords: convolutional neural network, skin segmentation, Indian sign language

1. INTRODUCTION

Humans are living in a society as a community because they are very similar when comes to the way of living, way of communicating, way of expressing etc. But after a glance on people, it was found that some are not as privileged as all of the others. Those are deaf and dumb people. These people have a large gap with normal speaking people under some circumstances such as communication and expression.

Various works have been using the technologies that are available but still there has not been any work which may revive the communication gap among people. There are several models made in different part of regions according to the requirement of the people such as American sign language (ASL), Arabic sign language, British sign language(BSL), Turkish sign language (TSL)etc. Indian sign language also sometimes called as Indo- Pak sign language have also been developed but there is much limitations to it.

Problem statement:

Not a proper system has been developed in order to resolve the problem of people with speaking disorder and deaf. Also multiple videos are available on websites that are not too helpful and moreover there is very less dual system (sign to speech and vice versa) developed.

Objective:

a) To aim towards developing a proper Indian sign language with all sets of alphabets and numbers.
b) To implement dual mode of communication. The spoken words will be taken as input and signs will be given as output.
c) To investigate different machine learning techniques like Support Vector Machines (SVM), Logistic Regression, K-nearest neighbors (KNN) and a neural network technique Convolution Neural Networks (CNN) for detection of sign language.

Proposed system:
A perfectly working model idea for Indian sign language, that includes all the alphabets and digits. It also includes dual mode of output that is, sign language converted into text/speech and speech converted in sign.

2. RELATED WORK

Deaf Mute Communication Interpreter- A Review [1]: This paper aims to cover the various prevailing methods of deaf-mute communication interpreter system. The two broad classification of the communication methodologies used by the deaf –mute people are - Wearable Communication Device and Online Learning System. Under Wearable communication method, there are Glove based system, Keypad method and Handicom Touch-screen. All the above mentioned three sub-divided methods make use of various sensors, accelerometer, a suitable micro-controller, a text to speech conversion module, a keypad and a touch-screen. The need for an external device to interpret the message between a deaf –mute and non-deaf-mute people can be overcome by the second method i.e online learning system. The Online Learning System has different methods. The five subdivided methods are- SLIM module, TESSA, Wi-See Technology, SWI_PELE System and Web-Sign Technology.

B. Hisham and A. Hamouda, "Supervised learning classifiers for Arabic gestures recognition using Kinect V2," SN Applied Sciences [2]**:** introduces a dynamic Arabic Sign Language recognition system using Microsoft Kinect which depends on two machine learning algorithms. However, Arabic sign language with this recent CNN approach has been unprecedented in the research domain of sign language. Therefore, this work aims at developing a vision-based system by applying CNN for the recognition of Arabic hand sign-based letters and translating them into Arabic speech. A dataset with 100 images in the training set and 25 images in the test set for each hand sign is also created for 31 letters of Arabic sign language. The suggested system is tested by combining hyperparameters differently to obtain the optimal outcomes with the least training time.

O. K. Oyedotun and A. Khashman, "Deep learning in vision-based static hand gesture recognition," Neural Computing and Applications [3]**:** who used CNN along with Stacked Denoising Autoencoder (SDAE) for recognizing 24 hand gestures of the American Sign Language (ASL) gotten through a public database. On the other hand, the proposal to use Convolutional Neural Network (CNN) for recognizing the Italian sign language was made by Pigou et al. Whereas Hu et al. had made a proposal for the architecture of hybrid CNN and RNN to capture the temporal properties perfectly for the electromyogram signal which solves the problem of gesture recognition. An incredible CNN model that automatically recognizes the digits based on hand signs and speaks the particular result in Bangla language is explained in, which is followed in this work. In as well, there is a proposal of using transfer learning on data collected from several users, while exploiting the use of deep-learning algorithm to learn discriminant characteristics found from large datasets.

3. SYSTEM DIAGRAM AND WORKFLOW

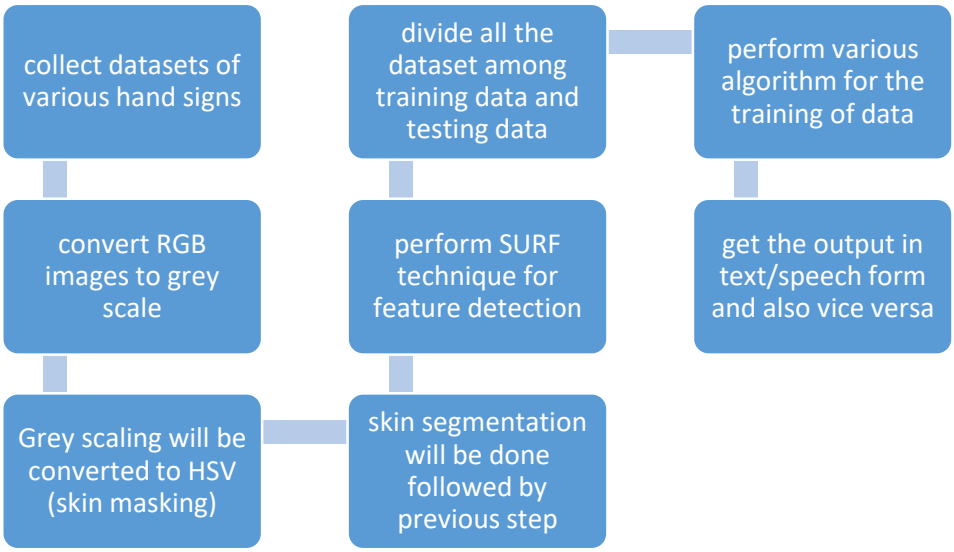

Fig 1: working of system

The above fig 1 refers to the working of the system. At beginning the dataset will be collected. Those datasets will go under various processes before getting into CNN layers. The processes which are required to go through before feeding the image to CNN layer are as follows.

a) PRE-PROCESSING OF IMAGE

Grey scaling:

Each pixel contains a luminance value, regardless of its color. Luminance for images can be characterized as brightness or intensity, which can be measured on a scale from black to white. So RGB form of image with be converted to grey scale.

Segmentation:

The main purpose of the segmentation phase is to remove the background and noises, leaving only the required area of image, which is the only useful information for the process. This is achieved via skin masking defining the threshold on RGB schema and then converting RGB colour space to grey scale image. Finally, canny edge technique is employed to identify and detect the presence of sharp discontinuities in an image, thereby detecting the edges of the figure in focus.

Feature Detection and Extraction:

The Speeded Up Robust Feature (SURF) technique is used to extract descriptors from the segmented hand gesture images as shown in fig 2. It is a fast and robust algorithm for local, similarity invariant representation and comparison of images. It is composed of three steps called feature extraction, feature description and feature matching.

Fig 2: SURF image

b) FEEDING THE IMAGE TO CNN LAYER

The image will be feed to CNN layers. CNN take an input image, assign relevance (learnable weights and biases) to various aspects/objects in the image, and distinguish between them. CNN has mainly 3 layers: convolutional layer, pooling layer and fully connected as shown below in fig 3.

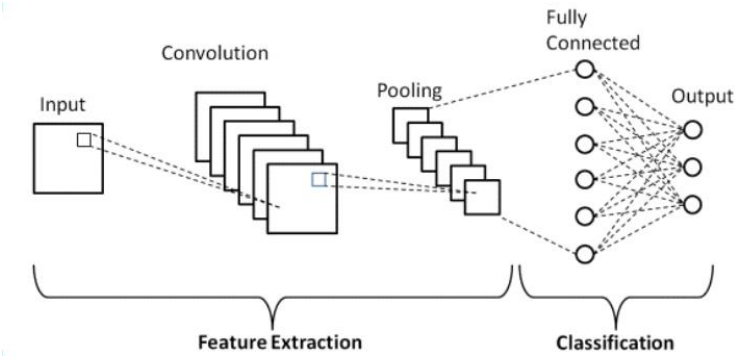

Fig 3: CNN basic architecture

c) CLASSIFICATION

For training the data set the required algorithms and classifiers as mentioned below. The classifiers that has to be used are as follows:

CNN: It is used for image recognition and classification that means to learn the position and scale of the image in different position. The convolutional layer captures feature like edges, color, gradient orientation etc. The pooling layer is used to perform a down sampling operation along the spatial dimension that is width and height, resulting in a fixed volume. Then locally connected layer accepts input from preceding layer, computes the class scores, and outputs a 1-D array with the same size as the number of classes.

KNN: KNN is used for high level of prediction. This is used for feature similarity that means new point is assigned a value based on how closely it resembles the points in training set. It is a famous method used for image classification. We use k=1 and Euclidean-distance to get smallest data process which is used for ranking.

Naïve Bayes classifier: Naïve Bayes classifier helps in building the fast machine learning model that can make quick predictions. It makes use of a language model to assign class labels to some instances, based on a set of features which can be numerically represented using statistical techniques.

Support vector machine: SVM is a set of supervised learning methods that are used for classification, regression and outlier's detection. Here all data items are plotted as a point in n dimensional plotting where n is the quantity of highlights with the value of every component being the value of that specific coordinator. At that point, classification is performed by finding hyperplane that isolates and separates the classes.

Logistic regression: logistic regression estimates the parameters of logistic model. It is used to understand the relationship dependent and independent variable.

Pattern matching and getting text

Training of the dataset will be done in the same way as explained above. When it is the turn of testing, an image of some sign using hand will be given to camera. If the pattern gets matched, it will soon give the output as text/speech and also vice versa.

4. PSEUDO CODE

```
# Preprocessing all the images to extract ROI i.e. hands

def preprocess_images(data,label):
    count=0
    for image in data:

        #reading image
        img=imageio.imread(image)

        #Converting image to grayscale
        gray_img=cv2.cvtColor(img,cv2.COLOR_RGB2GRAY)

        #Converting image to HSV format
        hsv_img=cv2.cvtColor(img,cv2.COLOR_RGB2HSV)

        #Defining boundary level for skin color in HSV
        skin_color_lower= np.array([0,40,30],np.uint8)
        skin_color_upper= np.array([43,255,255],np.uint8)

        #Producing mask
        skin_mask=cv2.inRange(hsv_img,skin_color_lower,skin_color_upper)

        #Removing Noise from mask
        skin_mask=cv2.medianBlur(skin_mask,5)
        skin_mask=cv2.addWeighted(skin_mask,0.5,skin_mask,0.5,0.0)

        #Applying Morphological operations
        kernel=np.ones((5,5),np.uint8)
        skin_mask=cv2.morphologyEx(skin_mask,cv2.MORPH_CLOSE,kernel)

        #Extracting hand by applying mask
        hand=cv2.bitwise_and(gray_img,gray_img,mask=skin_mask)

        #Get edges by Canny edge detection
        canny=cv2.Canny(hand,60,60)

        #save preprocessed images
```

```
path='ISL Recognition/Preprocessed Images/'
final_path=path+label+str('/')+str(count)+str('.png')
cv2.imwrite(final_path,canny)
count+=1
```

6. RESULT AND DISCUSSION

Signs shown on camera as input gives the output in the form of text as well as speech using the python text to speech conversion library. Reverse Sign Recognition is also implemented. The spoken word is taken as input and the corresponding sign images are shown in sequence. Google speech API is used for this purpose.

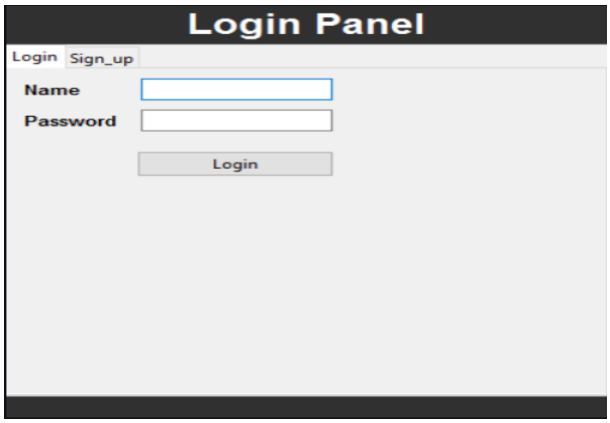

Fig 4: Login box

In the beginning, a login box will be opened as shown above in fig 4. The login panel consist of login option with name and password area. For the new user sign up option is also provided.

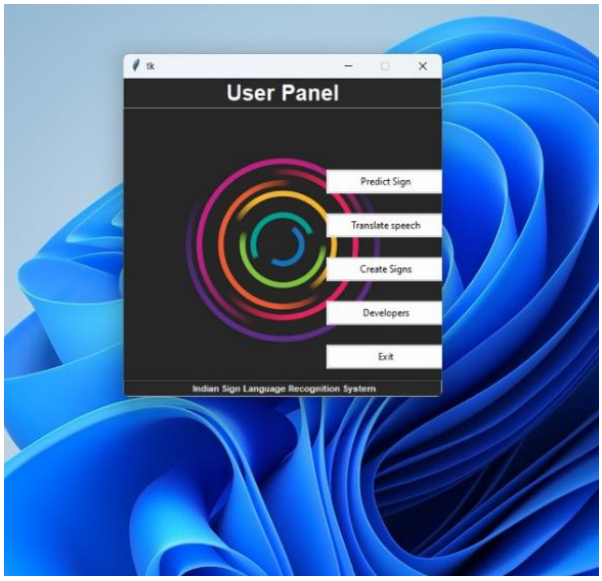

Fig 5: User box

Followed by login panel, there is a user panel that opens with multiple options as referred in fig no 5. The panel consists of 5 options namely predict sign, translate speech, create sign, developers and exit.

By selecting translate speech option from user panel (shown above), spoken word will be translated to sign. For testing, a word “open” is taken. When the word open is spoken, the sign of each letter of that word appears one by one on the screen. Here is the output of word shown below, consisting of fig no. 6,7,8 and 9.

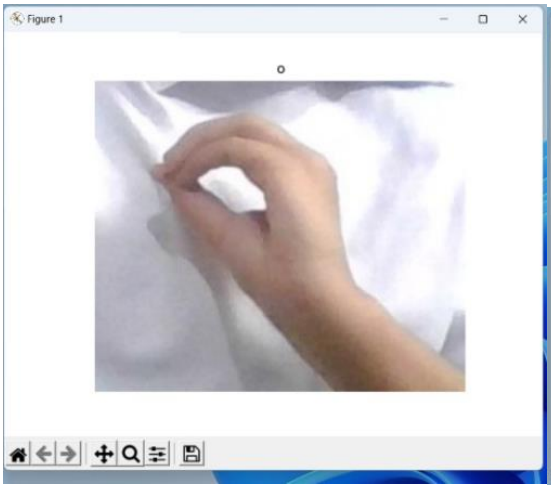

Fig 6: Alphabet O

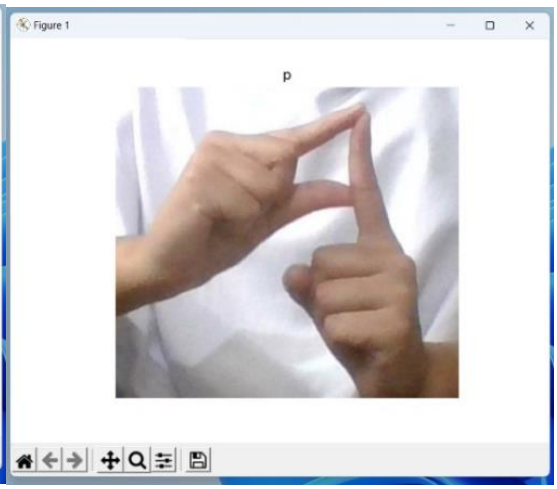

Fig 7: Alphabet P

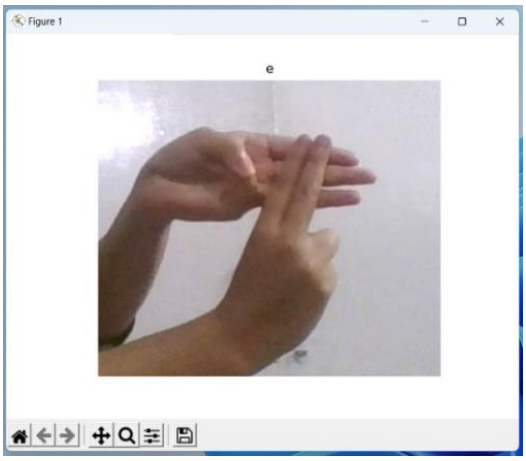

Fig 8: Alphabet E

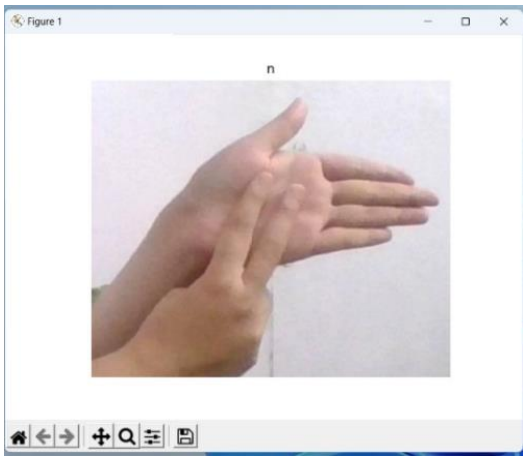

Fig 9: Alphabet N

7. CONCLUSION

This paper bridges the gap between normal people and deaf/dumb people by providing a new means of communication. It gives an opportunity to work and implement one's knowledge in order to help less fortunate using advanced technology. An attempt is made to work on a model for Indian sign language, that includes all the alphabets and digits. An incredible use of all the techniques is also done in order to provide the best and suitable information and implementation.

REFERENCES

1.Sanjay Meena. A Study of Hand Gesture Recognition Technique. Master Thesis, Department of Electronics and Communication Enginnering, National Institute of Technology, India, 2011

2. Olena Lomakina. Development of Effective Gesture Recognition System. TCSET'2012, Lviv-Slavske, Ukraine, February 21-24, 2012.

3. P. Subha Rajam and Dr. G. Balakrishnan. Real Time Indian Sign Language Recognition System to aid Deaf-dumb People. ICCT, IEEE, 2011.

4. Wilson A.D, Bobick A.F. Learning visual behavior for gesture analysis. In Proc. IEEE Symposium on Computer Vision, 2011

5. http://www.happinesspages.com/baby-sign-language-FAQ.html

6.http://mi.eng.cam.ac.uk/~cipolla/lectures/PartIB/old/IB-visualcodebook.pdf
7.https://github.com/shackenberg/Minimal-Bag-of-Visual-Words-Image-Classifier/blob/master/sift3.

8. http://en.wikipedia.org/wiki/YIQ

9. http://en.wikipedia.org/wiki/YUV