Sign Language Recognition(ISL)

Sofiya Mujawar

Professor-Computer Department
Department Of Computer Engineering
A.P.Shah Institute Of Technology
Thane(MH), India
ssmujawar@apsit.edu.in

Nidhi Singh
Department Of Computer Engineering
A.P.Shah Institute Of Technology
Thane(MH), India
singhnidhik2002@gmail.com

Ishanee Revankar
Department Of Computer Engineering
A.P.Shah Institute Of Technology
Thane(MH), India
ishaneeshet@gmail.com

Nidhi Heniya
Department Of Computer Engineering
A.P.Shah Institute Of Technology
Thane(MH), India
nrheniya02@gmail.com

Chirag Hegde
Department Of Computer Engineering
A.P.Shah Institute Of Technology
Thane(MH), India
hegdechirag321@gmail.com

Abstract— Sign language is the only way of communication for people with speaking and hearing inabilities. Understanding human emotions are considered a pattern recognition problem. Sign languages use hand gestures, facial expressions, and body movements for communication. There are many systems available for recognizing ASL that is American Sign Language, but not ISL, the sign language used in India. Our system recognizes Indian Sign Language and converts them into human-readable English text. Our focus is on recognizing Alphanumeric (0 to 9 and A to Z) signs/gestures by training the machines with static images. We have created the dataset containing 26 signs of A-Z letters and 10 signs of 0-9 numbers.

The proposed method stated in this paper is digital image processing techniques and neural networks for recognizing different signs. Only hand movements are considered for sign language interpretation. Mainly steps involved in recognition are Real-time image capturing through a web camera, image pre-processing, feature extraction, and classification. The model used is CNN for gesture recognition and ANN for classification. The model classifies images of all signs with an accuracy of 93% in presence of proper sunlight and white background

Keywords—Convolutional Neural Network, Indian Sign Language ,Deep Learning , Image processing , Image Classification.

I. INTRODUCTION

Language is a bridge of communication used by two beings. In a world this large there are thousands of languages being spoken. Considering India itself, we have a near about 117 different languages. But these are verbal languages, there comes languages other than verbal languages which deaf, dumb, blind people use to communicate. Sign language is one of them. Sign Language is used by deaf and dumb people. Sign languages comes under several Spoken languages such as English, Chinese, Japanese, Indian, etc. Sign Languages are non-verbal form of communication which uses visual sign or gestures to form a way of communication. Sign Language used by Indian people is called as Indian Sign Language i.e., ISL.

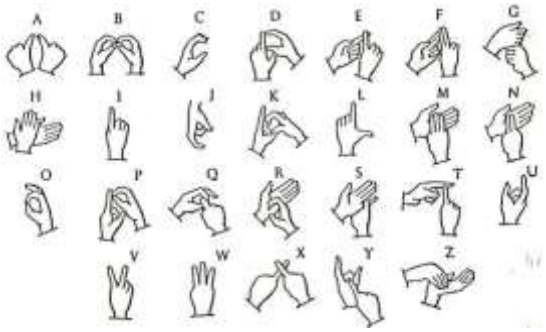


Figure 1 Alphabets in ISL

Figure 2 Numbers in IS;

Outside India, Sign languages have gained immense importance, but in India it hasn't yet snatched the spotlight. Because of the same, not many plans/projects have been created in/for ISL. But it is equally important to connect ISL to trendy technology too.

Sign languages are used as a primary means of communication by deaf and hard of hearing people worldwide. It is the most potent and effective way to bridge the communication gap and social interaction between them and the able people. Sign language interpreters help solve the communication gap with the hearing impaired by translating sign language into spoken words and vice versa. However, the challenges of employing interpreters are the flexible structure of sign languages combined with insufficient numbers of expert sign language interpreters across the globe. According to the World Federation of Deaf, more than 300 sign languages are used by more than 70 million worldwide. Therefore, the need for a technology-based system that can complement conventional sign language interpreters.

This project helps mute-deaf people and people in general to communicate with each other with ease. Basic aim of this project is to detect the signs gestured by a person and convert it to English text. Considering ISL is a little complicated since both hands are used to gesture a sign, 0-9 numbers will be detected using this project.

For this, we have built a project on ISL called as SIGN LANGUAGE RECOGNITION. In this project, gestures signed by a person are captured and recognized, thus detected signs are converted to readable English language and displayed on screen. For recognizing the signs, the Regions of Interest (ROI) are identified and tracked using the skin segmentation feature of OpenCV. The training and prediction of hand gestures are performed by applying CNN machine learning algorithm.

II. Literature Review

For the past decades, research on sign language recognition has been explored.

The initial process is to extract features from the frame sequences.[1] This will result in a representation, consisting of one or more feature vectors, also called descriptors. The second step is the classification using an artificial neural network .In [1] Temporal segmentation is used to predict the beginning and end frames of every gesture in the video sample. After preparing the dataset, the hand segmentation is done using a color segmentation method known as YCbCr color space. The image is converted to white and black by setting some constraints using Cb,Cr and Y values. Shape feature is derived from this binary image using distance transformation .Row and column projector vectors are calculated from the distance transformed image. Using Fourier Descriptors the hand shape is represented.

The classification tool used in [2] is Artificial Neural Network . A feed forward neural network in combination with a supervised learning method is used. Accuracy achieved is 91% with less computation time.

The feature selected in the paper [3] to reduce the recognition time is the number of finger tips in the image. Thinning is performed on image using distance transform and corner points are found using Harris corner detection algorithm. PCA is used for high dimensional data analysis (live data frames). Accuracy achieved in recognition through the above method is 94% .

After the acquisition of images in [4], they are processed using Canny Edge Detection technique to extract edges of the palm. Various studies convey the difficulties in recognizing bare hands of the signer used to portray the signs due to noise and other hindrances .The paper[5] presents the results obtained after retraining and testing the gesture dataset on a convolutional neural network model using Inception v3. The developed model consists of multiple convolutional filters .

In [6], the approach taken by Balbin et al was to use colored gloves for hands. For recognizing hand gestures, input images underwent various image processing methods . Initially, input images are converted into grayscale and a median filter is applied to denoise the image, the hand feature is detected and isolated from the background the model used is kohen self-organizing maps to identify patterns and groups dataset. The test accuracy of the above model is approximately 97.6%

In the study[7] they have compared the results of artificial neural networks ,support vector machines and hidden Markov based methods to recognize ASL Language.They collected data from Mayo armbands. A neural network was constructed with 26 features and 13 classes using a feed-forward network .It was a fully connected network.

The comparison was based on the recognition accuracy. The individual instance classification was used for ANN and SVM while the grouped instance classification was used for the HMM. The ANN method yielded an overall accuracy of 93.79% in terms of correctly recognized test instances and the best recognition accuracy result for SVM was 89.05% .The best performance of the HMM technique is reported at 85.90% . Referring to various conducted experiments, it was concluded that ANN has a better performance compared to other techniques

The research done in paper[8] is based on recognizing American sign language that includes gesture and alphabet recognition. The researchers have used convolutional neural networks for classification of images and keras for training. The proposed system uses skin-color modeling technique,a skin color range is explicitly defined in order to differentiate the hand pixels from the pixels of the background .The accuracy achieved through this method was around 93.76% , 90.04% accuracy was achieved by ASL alphabet recognition ,97.52% for static word recognition. According to the study SLR architecture is based on two different ways of inputs: data glove based and vision based .Using smart gloves it is easier to get the position of hands and the orientation but has limited movement . Vision based sign language recognition involves the feature extraction step, in this method image is given to the system and images are processed and features are extracted from it by image processing and computer vision method, then the recognizer learns from the patterns of the dataset of image and through machine learning algorithm it recognizes the image.

This paper uses multiple image processing methods to recognize hand gestures. Initially, input images are converted into grayscale and a median filter is applied to denoise the image, the hand feature is detected and isolated from the background the model used is kohen self-organizing maps to identify patterns and groups dataset. The test accuracy of the above model is approximately 97.6%

In the study[7] they have compared the results of artificial neural networks , support vector machines and hidden

Markov-based methods to recognize ASL Language. They collected data from Mayo armbands. A neural network was constructed with 26 features and 13 classes using a feed-forward network .It was a fully connected network.
SVM is a binary classification technique. The comparison was based on the recognition accuracy. The individual instance classification was used for ANN and SVM while the grouped instance classification was used for the HMM. The ANN method yielded an overall accuracy of 93.79% in terms of correctly recognized test instances and the best recognition accuracy result for SVM was 89.05% .The best performance of the HMM technique is reported at 85.90% . Referring to various conducted experiments, it was concluded that ANN has a better performance compared to other techniques.

In the[8] proposed system vision based approach is deployed , a web camera is used to obtain data from the signer . They converted some of the letters in ASL alphabet so that it does not affect the accuracy of the system .Thirty five words were chosen and divided into four categories: family, communication ,and transportation. The methods used in this paper are gathering training data. That was done by capturing images using python . Images were then converted into black and white then skin color was detected by using cv2.cvtColor. The convolutional layer used 16 filters with 2*2 kernel. The resulting system was tested by thirty individuals.
The study in this paper is automatic sign language recognition based on a random forest machine learning algorithm.
.The system uses a low level visual features to recognize the sign language and achieved an accuracy of about 95%.
According to study, each sign consists of phonemes such as hand -shapes ,locations and movements that are made using one or both the hands . This paper uses the idea that signs can be broken into phonemes and then processed . This system uses a vision based method for sign detection. The methods used in the system are image capturing , skin detection ,feature extraction , modeling and identification of signs . The cameras are used to take the images and skin color is used to detect the hands. The problem in this case is that the perfect skin color range may not be the same for all the captured video and some objects in the background may have the same color as those of the hands. For the purpose of feature extraction the hand shapes are encoded using seven hu-moments and movements as XORs of two consecutive location grids . To encode the hand location they have used a 10 by 10 matrix with the center of the face used as reference, Viola jones face detector is used to detect the face. The size of the face is determined using data and implementation provided in Opencv library. In this study they have mentioned that ideal sign language is independent of context, content and vocabulary and robust with regard to signer identity.
To encode the types of body movements they compared the locations of hands/face in the current frame with respect to the previous frame. The system is trained on 50% of the dataset and tested on the other 50% of dataset. The system was trained randomly on clips of eleven signers and tested on clips of eight singers. The model is evaluated in terms of precision, recall and F1 score and achieves an accuracy of 95% .
The research work in this system[8] aims at developing an automatic recognition system for Indian sign language numerals . They have used their own database that consists of 1000 images, 100 images representing every sign. The system uses Shape descriptors ,Scale invariant feature transform and)and Histogram of Oriented Gradients (HOG) techniques are used for extracting desired features.

III. Problem Statement

Language is the only medium by which we can share our thoughts and communicate but for a person with disability faces difficulty in communicating. This people communicate using sign language. But sign language is not understood by everyone, because of this a deaf person and someone with speaking disability are not able to stand in race with others and fail to communicate. This project aims to implement a system that recognizes Indian Sign Language and converts it to readable English text in real time, making it easier for such people to communicate .

IV. Methodology

V. The system is a vision based approach. All the signs are represented with bare hands and so it eliminates the problem of using any artificial devices for interaction.

A] Dataset Generation:

We have created our own dataset with 250 images for each sign representing 0-9 numbers and 250 images for each sign representing A-Z numbers.

Fig : signs of numbers

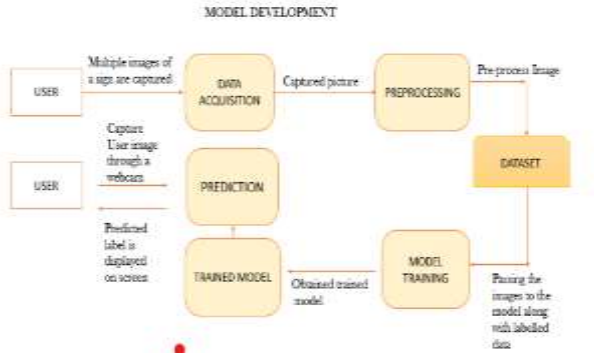
Step1:OpenCV library is used to capture the images for our dataset. We captured 120 images for each symbol ,for the train set and around 12 images for the test set.

Step 2:We capture the frame shown by the webcam. In each frame we have defined ROI(region of interest) denoted by blue rectangle

Step 3:We extract the frame inside the ROI ,which is in RGB,.later we convert it into grayscale.

Step 4:Apply threshold to the captured frame to get processed image
Processed image is passed through the CNN model

B] Implementation



C] CNN Model:

It is most widely used for image classification. In the layers of CNN, the neurons are arranged in 3 dimensions width, height, and depth. The neurons in a layer will only be connected to a small region of the layer before it, instead of all of the neurons in a fully-connected manner. Moreover, the final output layer would have dimensions that represent the number of classes, because by the end of the CNN architecture we will reduce the full image into a single vector of class scores.

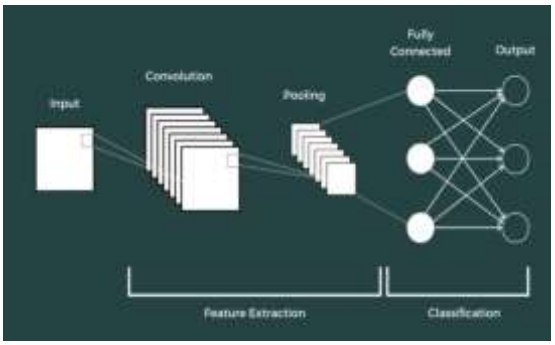


Fig CNN Architecture

1. Convolution layer:

The first layer to extract information from the images in a CNN is the convolutional layer

2. Strides:
Stride is the number of pixels shifted over the input matrix. When the stride is 1 then the filters are moved 1 pixel at a time. When the stride is 2 then we move the filters to 2 pixels at a time.

3. Activation function:

The main purpose of Activation function is to introduce non-linearity into the output of neuron. Activation function makes the back propagation possible

ReLU(Rectified linear unit):
ReLU introduces non-linearity.It is implemented in hidden layers of Neural Network.
It computes the function $f(\kappa)=\max(0,\kappa)$.It is basically the threshold at zero.
ReLU is better in performance in comparison to tanh or sigmoid.

4. Pooling Layer :

The pooling layer is used to decrease the size of the activation matrix and this, in turn, reduces the learnable parameters.

It helps in reducing the spatial size of the representation, which decreases the required amount of computation and weights. We have used Max Pooling which takes only maximum values from a chosen window size.

5. Fully connected layer:
Neurons in this layer have full connectivity with all neurons in the preceding and succeeding layer. This is why it can be computed as usual by a matrix multiplication followed by a bias effect. The FC layer helps to map the representation between the input and the output.

Softmax Function: We have used this activation function at the end of the model definition. It limits the output into a range from 0-1 which represents the probability of output of a possible class. It gives output as a vector containing the probabilities of each possible outcome

D] Training:

The dataset is split into training and testing dataset, the dataset is passed to convolutional neural network model

E] Testing:

Once the model is trained, it is tested with images of number and alphabet in order to get the accuracy of the model. Captured image is passed to loaded model as an input to the convolutional neural network graph that gives a confidence level to the class of an image.

This system was tested by 10 individuals and each of them got all the signs correctly recognized. Initially the accuracy of this system was about 75%,

After increasing the size of dataset, we were able to achieve an accuracy of about 93%.

Software Requirments :

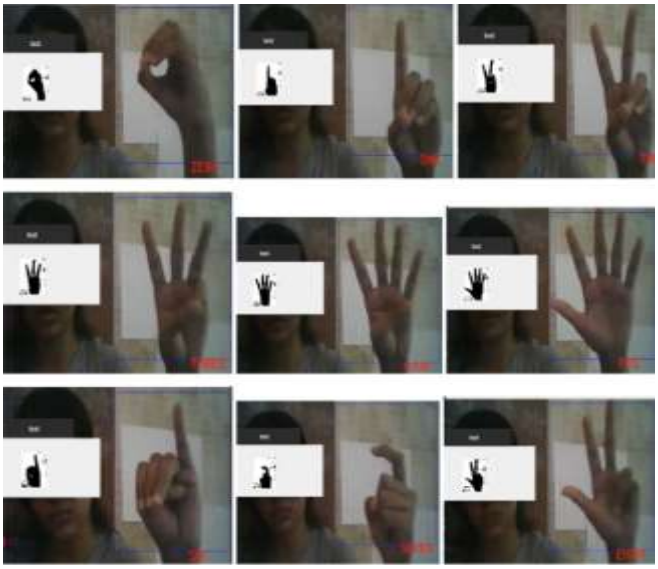
1. Python: General Programming Language used to code the model, includes several libraries.
2. Python libraries: Numpy, scipy,keras,tensorflow, PIL (Python Imaging Library), pandas, os.
3. VS Code/ Anaconda: IDE to run, train and test the ML model. Plenty of extensions, open-source, cross-platform support,

VI. Result

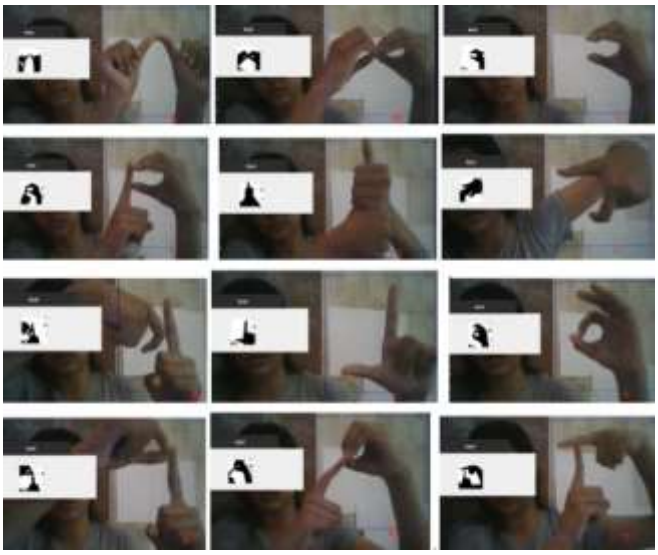
F] Prediction:

We designed our system to predict the label for every sign language depicted in the image.
The system correctly recognized every sign language with an accuracy of 93%.
There were some alphabets like E F H I MN R that we excluded from our dataset.
Upon execution a region of interest is provided in the Open cv frame is provided to user so that only user's hand gesture is captured , the captured image gets converted into black and white and is fed to convolutional neural network model for classification. The predicted value is displayed on the screen below the region of interest.

Prediction of ISL numbers



Results of prediction of ISL alphabets



VII. Conclusion

This project successfully detects the signs and converts it to text. Overall our study revealed that it is necessary to have a white background to obtain high accuracy in recognition.

In future this will be extended to detecting sign language as whole that is including alphabets, greetings and complete sentences. We are planning to make this system dual-side useful, that is along with abled people, disabled people too would be able to use it. Guides to ISL and videos for basic greeting and useful signs will be included.

References

[1] Lionel Pigou (B), Sander Dieleman, Pieter-Jan Kindermans, and Benjamin Schrauwen ELIS, "Sign Language Recognition Using Convolutional Neural Networks", Ghent University, Ghent, Belgium

[2] Adithya V., Vinod P. R., Usha Gopalakrishnan," Artificial Neural Network Based Method for Indian Sign Language Recognition", Proceedings of 2013 IEEE Conference on Information and Communication Technologies (ICT 2013)

[3] Divya Deora, Nikesh Bajaj," INDIAN SIGN LANGUAGE RECOGNITION", 2012 1st International Conference on Emerging Technology Trends in Electronics, Communication and Networking

[4] P. Subha Rajam Dr. G. Balakrishnan, "Real Time Indian Sign Language Recognition System to aid Deaf-dumb People", J.J. College of Engineering & Technology, Trichy, Tamil Nādu, India.

[5] Das, Aditya; Gawde, Shantanu; Suratwala, Khyati; Kalbande, Dhananjay (2018). [IEEE 2018 International Conference on Smart City and Emerging Technology (ICSCET)-Mumbai (2018.1.5-2018.1.5)] 2018 International Conference on Smart City and Emerging Technology (ICSCET) - Sign Language Recognition Using Deep Learning on Custom Processed Static Gesture Images.

[6] J. R. Balbin, D. A. Padilla, F. S. Caluyo, J. C. Fausto, C. C. Hortinela, C. O. Manlises, C. K. S. Bernardino, E. G. Finones, and L. T. Ventura, "Sign language word translator using neural networks for the aurally impaired as a tool for communication," in Proc. 2016 6th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), 2016, pp. 425-429

1) [7] Fatmi, Rabeet; Rashad, Sherif; Integlia, Ryan (2019). [IEEE 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC) - Las Vegas, NV, USA (2019.1.7-2019.1.9)] 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC) - Comparing ANN, SVM, and HMM based Machine Learning Methods for American Sign Language Recognition using Wearable Motion Sensors.

[8] Tolentino, Lean Karlo S., et al. "Static sign language recognition using deep learning." International Journal of Machine Learning and Computing 9.6 (2019): 821-827. Vol. 9, No. 6, December 2019