EasyChair Preprint

№ 7355

Indian Sign Language Interpreter

Aayush Shukla, Chetan Parmar and Ashwini Deshmukh

January 21, 2022

INDIAN SIGN LANGUAGE INTERPRETER

Aayush Shukla
Shah & Anchor Kutchhi Engineering College
Chembur, India
aayush.shukla@sakec.ac.in

Chetan Parmar
Shah & Anchor Kutchhi Engineering College
Chembur, India
Chetan.parmar@sakec.ac.in

Ashwini Deshmukh
Shah & Anchor Kutchhi Engineering College
Chembur, India
ashwini.deshmukh@sakec.ac.in

***Abstract*— The sign language is a very important way of communication for mute people. In sign language each gesture has a specific meaning. So therefore complex meanings can be explain by the help of combination of various basic elements. Sign language is a gesture based language for communication of mute people. It is basically a non-verbal language which is usually used to mute people to communicate more effectively with each other or normal people. Sign language contains special rules and grammar's for expressing effectively. Basically there are two main sign language recognition approaches image- based and sensor-based. Gesture recognition is gaining importance in many applications areas such as human interface, communication, multimedia and security. Typically Sign recognition is related as image understanding. It contains two phases: sign detection and sign recognition. Sign detection is an extracting feature of certain object with respect to certain parameters. Sign recognition is recognizing a certain shape that differentiates the object from the remaining shapes. Language, especially in the cases when no alternative communication is available.**

I. INTRODUCTION

Special abled people (Deaf) have used sign languages throughout history. One of the earliest written records ofa sign language is from the fifth century BC, in Plato's Cratylus, where Socrates says: "If we hadn't a voice or a tongue, and wanted to express things to one another, wouldn't we try to make signs by moving our hands, head, and the rest of our body, just as dumb people do at present?" Sign languages do not have any linguistic relation to the spoken languages of the lands in which they arise. The correlation between sign and spoken languages is complex and varies depending on the country more than the spoken language. For example, Australia, Canada, New Zealand, the UK, and the US all have English as their dominant language, but American Sign Language (ASL), used in the US and English-speaking Canada, is derived from French Sign Language while the other three countries use varieties of British, Australian and New Zealand Sign Language, which is unrelated to ASL. Indo-Pakistani Sign Language (IPSL) is the predominant sign language in South Asia, used by at least several hundred thousand deaf signers. The Deaf communities of the Indian subcontinent are still struggling for IPSL to gain the status of sign language as a minority language. Though sign language is used by many deaf people in the Indian subcontinent, it is not used officially in schools for teaching purposes. In 2005, the National Curricular Framework (NCF) gave some degree of legitimacy to sign language education, by hinting that sign languages may qualify as an optional third language choice for hearing impaired students. NCERT in March 2006 published a chapter on sign language in a class III textbook, emphasizing the fact that it is a language like any other and is "yet another mode of communication." The aim was to create healthy attitudes towards the differently abled. Although discussion of sign languages and the lives of deaf people is extremely rare in the history of South Asian literature, there are a few references to deaf people and gestural communication in texts dating from antiquity. Symbolic hand gestures known as mudras have been employed in religious contexts in Hinduism, Buddhism and Zoroastrianism for many centuries, although these religious traditions have often excluded deaf people from participation in a ritual or religious membership. In addition, classical Indian dance and theatre often employs stylized hand gestures with particular meanings

Sign language is a visual language and consists of 3 major components:

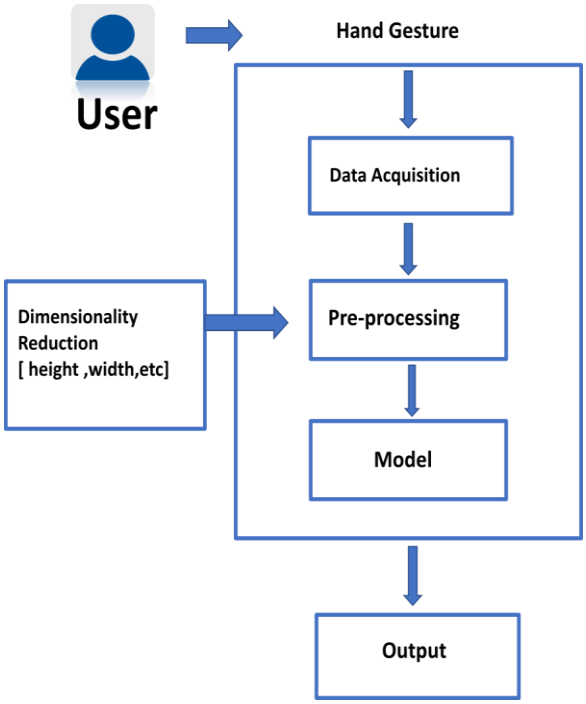
- Fingerspelling: used to spell words letter by letter.
- Word level sign vocabulary: Used for the majority of communication.
- Non-manual features: Facial expression and tongue, mouth, and body position.

This project is primarily based on First part of the sign language Fingerspelling

II. PPROBLEM STATEMENT

A Normal person if visits a mute person and try to communicate with him, will face difficulties in trying to understand what a mute person is trying to express. These kind of people cannot spend normal life. They face a lot communication issues at every point. Also they get certain boundaries and limitations to their goals and professional aims. Hence they get demotivated and Inferiority Complex. A Sign Language Interpreter is needed for such type of people to ease their life.

III. FLOW CHART



IV. STEPS

The steps to proceed with these are:

- Data Extraction
- Data pre-processing
- Feature recognition using neural network model

1.Data Extraction

- Using computer vision

Computer vision tasks include methods for acquiring, processing, analyzing and understanding digital images, and extraction of high-dimensional data from the real world in order to produce numerical or symbolic information, e.g., in the forms of decisions. Moreover, this method is more effective as it does not require overhead cost for equipment, but image processing machine like mobile, laptop, computer, etc. The complication here will be due to large variation on surrounding and noise removal from region of Interest

In order to make a dataset we used computer vision in form of camera sight as primary data collection tool

Different types present were:

- Infrared camera
- Ultraviolet light sensor
- Visible light camera

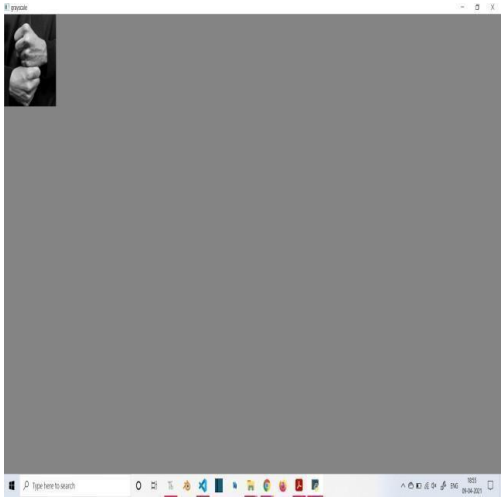
And more but to make it affordable and easy to use we stick to visible range light capturing camera

2. Data pre-processing:

- Extract the region of Interest as whole of camera view captures most of the noisesurrounding which are redundant thus cropping image to a square shaped window.

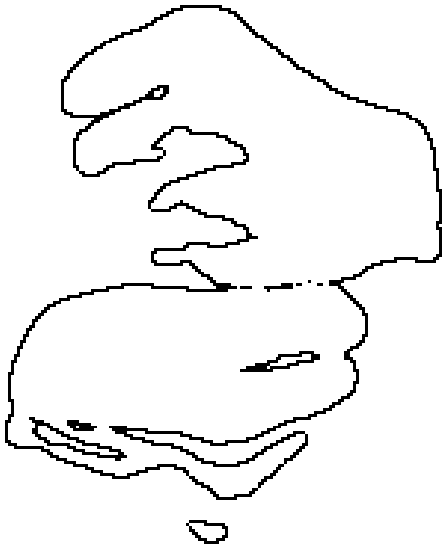
- We first of all converted the 3-channel image (RBG) into grayscale in order todecrease the operation time and complexity of processing.

- We resized the image into 200x200 pixel image so as to further more reduce the processing time and space complexity of the algorithm.



- We introduced a gaussian filter in order to reduce the noise in image.

- Now to extract the edge map we apply adaptive threshold filter of gaussian nature



3. Feature recognition/ extraction using neural network model:

Algorithm Layer:

- Apply gaussian blur filterandthreshold to the frame taken with OpenCV to get the processed image after feature extraction.

- This processed image is passed to the CNN model for prediction and if a letter is detected for more than 25 frames then the letter is printed and taken into considerationfor forming the word.

- Space between the words is considered using the blank symbol which is an empty frame with no sign.

CNN Model:

- **1st Convolution Layer:** The input picture has resolution of 200x200 pixels. It is first processed in the first convolutional layer using 64 filter weights.

- **1st Pooling Layer:** The pictures are down sampled using max pooling of 3x3 i.e., we keep the highest value in the 3x3 square of array. Therefore, our picture is down sampled.

- **2nd Convolution Layer:** Now, this output of the first pooling layer is served as an input to the second convolutional layer. It is processed in the second convolutional layer using 128 filter weights (2x2 pixels each).

- **2nd Pooling Layer:** The resulting images are down sampled again using max pool of 3x3 and is reduced to even lesser resolution of image.

- **3rd Convolution Layer:** convolutional layer using 256 filter weights (2x2 pixels each).

- **3rd Pooling Layer:** The resulting images are down sampled again using max pool of 3x3 and is reduced to even lesser resolution of image.

- **Flatten Layer:** It is used to convert the 2D pixel array into linear form in order to produce converge it into 27 class of hand signs

- **Final layer:** The output of the 3rd Densely Connected Layer serves as an input forthe final layer which will have the number of neurons as the number of classes we are classifying (alphabets + blank symbol).

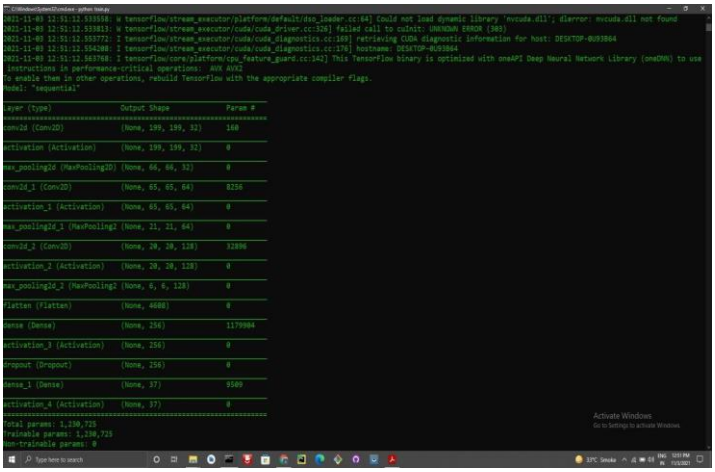
- **Activation Function:** We have used ReLu (Rectified Linear Unit) in each of the layers (convolutional as well as fully connected neurons). ReLu calculates max(x,0) for each input pixel. This adds nonlinearity to the formula and helps to learn more complicated features. It helps in removing the vanishing gradient problem and speeding up the training by reducing the computation time. At the last activation function, we used SOFTMAX function. It is used as the activation function in the

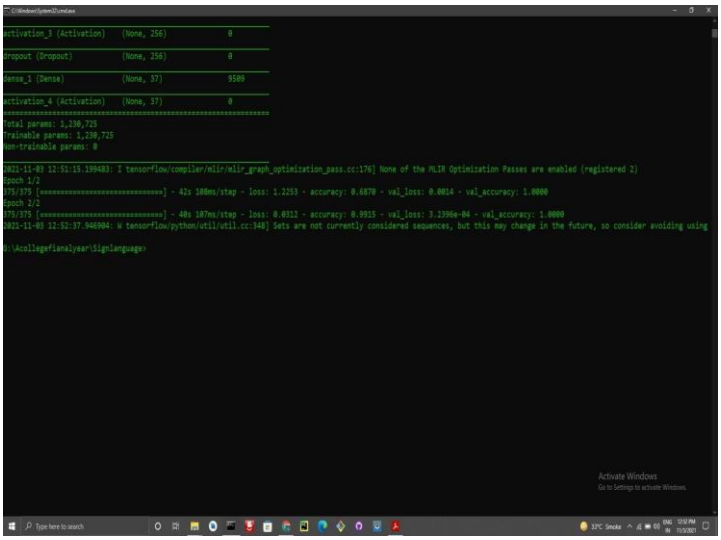
  output layer of neural network models that predict a multinomial probability distribution. That is, SoftMax is used as the activation function for multi-class classification problems where class membership is required on more than two class labels.

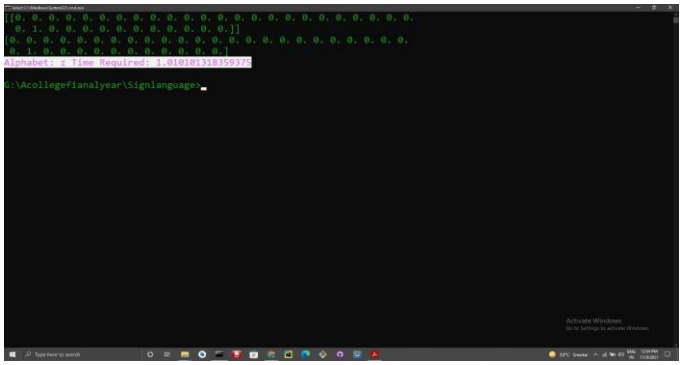
- **Pooling Layer:** We apply Max pooling to the input image with a pool size of (3, 3) with ReLu activation function. This reduces the amount of parameters thus lessening the computation cost and reduces overfitting.

- **Dropout Layers:** The problem of overfitting, where after training, the weights of the network are so tuned to the training examples they are given that the network doesn't perform well when given new examples. This layer "drops out" a random set of activations in that layer by setting them to zero. The network should be able to provide the right classification or output for a specific example even if some of the activations are dropped out.

- **Optimizer:** We have used Adam optimizer for updating the model in response to the output of the loss function. Adam combines the advantages of two extensions of two stochastic gradient descent algorithms namely adaptive gradient algorithm (ADA GRAD) and root mean square propagation (RMSProp).







V. RESULT

We convert our input images (RGB) into grayscale and apply gaussian blur to remove unnecessary noise. We apply adaptive threshold to extract our hand from the background and resize our images to200x200.

We feed the input images after pre-processing to our model for testing after applying all the operations mentioned above.
The prediction layer estimates how likely the image will fall under one of the classes.

VI. Conclusion

In this report, a functional vision based Indian sign language recognition for Mute people have been developed for ISL alphabets, numbers and basic gestures. This model could be used by especially abled person to be able to convey their hand sign or gesture language and aid an ordinary person to translate hand sign languagein order to make the communication more fluent between a person with speaking disability but has knowledge about Indian sign language. Most of the people are not able to communicate due to either the lack of knowledge of sign language to other peoples. Most of the programs present usually have only American signlanguage as their region of processing, but this application is specially designed for Indian Sign Language (ISL) Translation.

VII. Future Scope

- Addition of more basic gestures in model
- Achieve higher accuracy even in case of complex backgrounds by trying out various background subtraction algorithms.
- Improving the pre-processing to predict gestures in bright light conditionsand extremely low light conditions with a higher accuracy without using Black Background.

References

[1] https://www.kaggle.com/vaishnaviasonawane/indian-sign-language-dataset

[2] https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.pinterest.com%2Fpin%2F29568953176558472 8%2F&psig=AOvVaw2NTOz2r_JONtzVNTYxHsPQ&ust=1621322534718000&source=images&cd=vfe&ved=0CAIQjRxqFwoTCLC6hZ2X0PACFQAAAAAdAAAAABAD

.

[3] https://cse.iitk.ac.in/users/cs365/2015/_submissions/vinsam/report.pdf

[4] https://acadpubl.eu/jsi/2017-117-20-22/articles/20/2.pdf.

[5] https://www.ijrte.org/wpcontent/uploads/papers/v7i6/F2746037619.pdf

[6] https://www.sciencedirect.com/science/article/pii/S1877050918321367