ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 8, Issue 8, August 2021

Bi-Directional Sign Language Recognition Using Machine Learning

[1] Prajakta Bimalkhedkar, [2] Devika Rembhotkar, [3] Sanjana Pisute, [4] Mr. Jitendra Chavan

[1][2][3][4] Department of Information Technology, Marathwada Mitra Mandal's College of Engineering Pune, India

***Abstract*---- Most of the data in the computer world is available to a few people who can read or understand a specific language. People having hearing, visual or audio impairments face issues when communicating with others. We aim to create a technique with the help of which we can establish a sound communication system between normal and deaf/dumb and blind people. Sign-To-Speech-Text-To-Image (SSTI) is a technology that converts written text into a voice that can be understood by human beings and displays text. A computer based system is an SSTI synthesizer that can read any text that is provided by standard input devices. In particular, the Marathi Sign to Speech and Text to Speech conversion application is used for the localization of computer applications.**

***Keywords*— Stock Prediction, Data Analysis, Natural Language Processing, Machine Learning**

I. INTRODUCTION

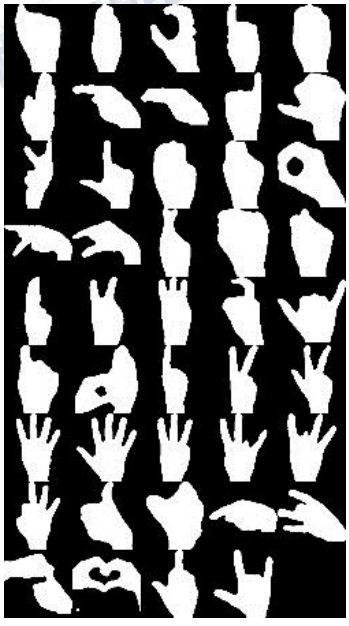
In the present age where a lot of applications and software are easily available for people to communicate, there is a distinct lack of the same for those having hearing/speech impairments. Sign Language is the evocative form of communication for those having speech and hearing impairments to communicate with normal person but a normal person generally is unable to understand sign language. So, to break this barrier of communication there needs to be a system that can enable conversion of sign language to voice or text and voice or text to sign language and do it in real time. There are such certain systems that exist that have good accuracy but they require external hardware like gloves which might not be convenient for all. We have designed this approach that consists of a communications system that is built using Machine Learning techniques such as Image Processing and Deep Learning to provide real-time conversion of sign language to text and text to sign language.

II. PROPOSED SYSTEM

The user can send the sign or text to the device here in this suggested system. We can train our system before that by supplying sign images with training datasets. Whenever we can provide the computer with a sign or text, pre-processing can be done on the image or text at that moment. Unwanted data can be extracted from the given image in image pre-processing. In the next stage, extraction of features can be carried out on the image, extraction of features is nothing but structure or characteristics of the image or text provided. After that, using the CNN algorithm, we can identify our image or text. And finally, we get the output of the given data such as Sing is converted into Speech and Text is converted into the image by performing all these operations. We can quickly identify signals or messages.

III. METHODOLOGY

We have created our own dataset using OpenCV Library. For each gesture, we have captured 2400 images which were 50X50 pixels. Initially we have added 26 alphabets from A to Z for the purpose of testing, ten integers ranging from 0 to 9, and 8 words. We have used the same dataset to convert sign into text and then text into sign.



ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 8, Issue 8, August 2021

A. Data Pre-processing

The 50X50 pixels images were converted to grayscale and all the images were flipped which led to the creation of 4800 images for each gesture. There were total of 2,11,200 images. Then the file 'gesture_db.db' was created. We used this database file to convert sign to text.

B. Sign to Text

1) Image Processing:

The image processing part is mainly done using with the help of Open CV library and Python language.

This phase composed of three steps:

i. *Noise Detection*: This is done using median filtering. To keep edges and reduce noise, median filtering is utilized. To determine if a pixel represents its surroundings or not, it is entangled with its neighbors.

ii. *Skin Detection*: It aims to find regions that have human faces and hands in images and to reject as much "non- skin" parts.

iii. *Skin Classifier and Binary classification*: Used to indicate the boundary of the skin color class in a feature space by identifying whether a pixel is part of the skin or not. All pixels in the image plane are classified into object and background pixels.

2) Image Recognition:

The image obtained after the image processing step is then given as an input to convolutional neural network which is then used to classify the images appropriately.

3) Image to Text Conversion:

The classified image is finally displayed according to what words are signed on the screen.

C. Text to Sign Conversion

i. Text Classification:

The text to be converted into sign was taken as input.

The text input was then parsed and after parsing the words left were searched in the database.

ii. Text to Image Conversion:

After searching in the database, if the corresponding sign is found, then the program returns the sign as the output to the user.

D. Text to Speech Conversion

Speech synthesis techniques are used to convert input Text to computer generated voice.A TTS Engine converts written text to a phonemic representation, then converts the phonemic representation to waveforms that can be output as sound.

IV. ARCHITECTURE

We have used Keras to implement the CNN Architecture which consisted of multiple convolutional and dense layers. Our model has total 9 layers. All the layers and their respective parameters are specified in figure 1.

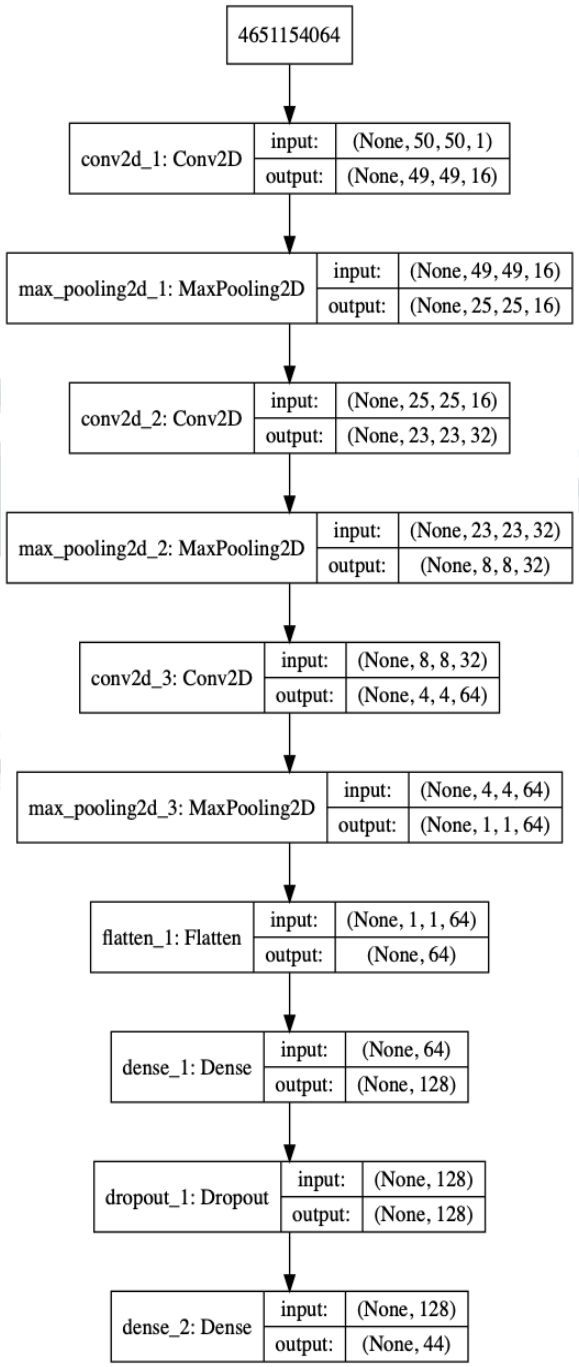


Figure 1: Layers of CNN Model

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 8, Issue 8, August 2021

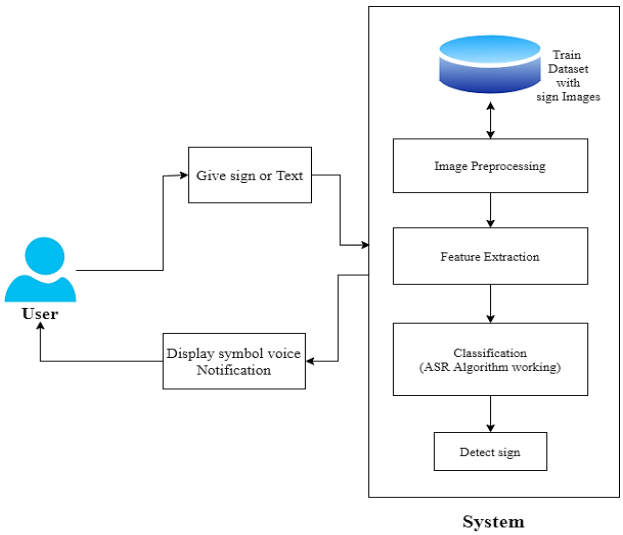


Figure 2: System Architecture

V. RESULTS

We were able to get significant results by using Machine Learning Algorithms. We were able to significantly minimize processing time by using two-dimensional images of 50X50 pixels.

1. Sign to Text Results

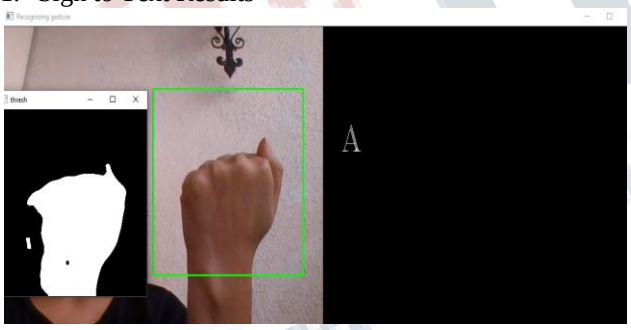


Figure 3: Output for A

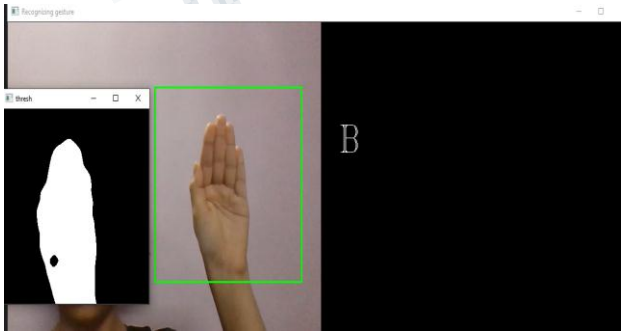


Figure 4: Output for B

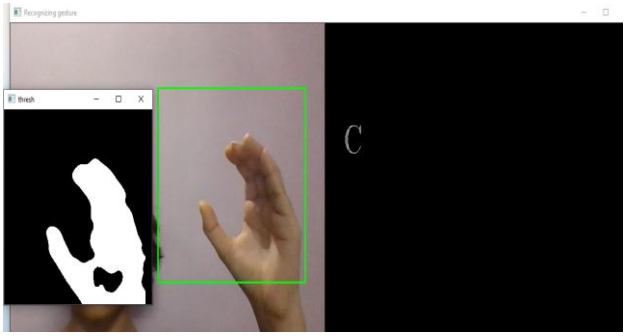


Figure 5 : Ouput for C

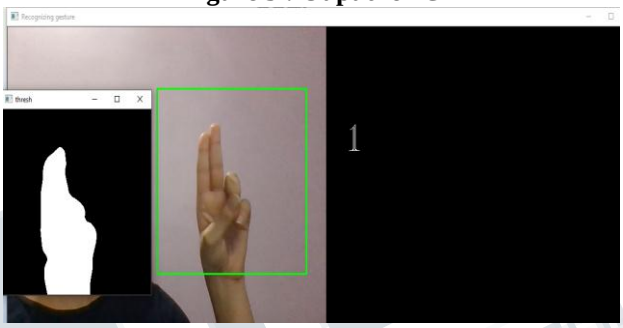


Figure 6: Output for 1

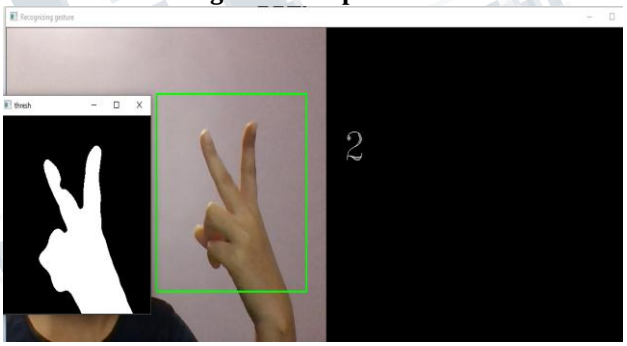


Figure 7: Output for 2

2. Text to Sign Results

Figure 1: GUI

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 8, Issue 8, August 2021

Figure 2 : Output for A

VI. PROBLEMS FACED

Sign languages are very broad and differ from country to country in terms of gestures, body language and face expressions. The grammars and structure of a sentence also varies a lot. In our study, learning and capturing the gestures was quite a challenge for us since the movement of hands had to be precise and on point. Some gestures are difficult to reproduce. And it was hard to keep our hands in exact same position when creating our dataset.

VII. FUTURE WORK

We look forward to use more commonly used words and phrases used in our datasets and improve the model so that it recognizes more alphabetical features while at the same time get a high accuracy. We would like to enhance the system by adding speech recognition so that blind people can benefit as well. The software can also be made open source so that developers from all around the world can make new advancements to it.

VIII. CONCLUSION

Many breakthroughs have been made in the field of artificial intelligence, machine learning and computer vision. They have immensely contributed in how we perceive things around us and improve the way in which we apply their techniques in our everyday lives. Many researches have been conducted on sign gesture recognition using different techniques like ANN, LSTM and 3D CNN. Sign recognition system is implemented to translate sign language gestures into the corresponding computer generated/human speech gestures for Blind/Normal to Deaf/Hard-of-Hearing. This system is divided into two parts. The first part is the static sign recognition in which alphabets, finger spellings or words that require no motion of hands or face are recognized. The second part is the voice recognition which is either isolated or continuous, isolated means where word by word is supplied to the server to recognize whereas continuous means a complete sentence is supplied to the server to recognize. This system can operate on a commodity PC with low-cost cameras.

REFERENCES

[1] Mariam Moustafa Reda, Nada Gamal Mohammed, Rania Ahmed Abdel Azeem Abul Seoud "SVBiComm: Sign-Voice Bidirectional Communication System for Normal,"Deaf/Dumb "and Blind People based on Machine Learning", 978-1-5386-4427-0/18 ©2018 IEEE.

[2] Tanuj Bohra,Shaunak Sompura,Purva Raut, Krish Parekh "Real-Time Two Way Communication System for Speech and Hearing Impaired Using Computer Vision and Deep Learning", Second International Conference on Smart Systems and Inventive Technology (ICSSIT 2019) IEEE Xplore Part Number: CFP19P17-ART; ISBN: 978-1-7281-2119-2.

[3] T .MeeraDevi, K . M .ShravanaRaju "Portable Communication Aid for Specially Challenged: Conversion of Hand Gestures into Voice and ViceVersa", 978-1-5386-9432-9/18 ©2018 IEEE.

[4] Mr.Rajesh M., Ms. Bindhu K. Rajan , Ajay Roy, Almaria Thomas K, Ancy Thomas, Bincy Tharakan T, Dinesh C "Text recognition and face detection aid for visually impaired person using RASPBERRY PI", IEEE Xplore.

[5] Abhishek Mathur, Akshada Pathare, Prerna Sharma, Sujata Oak, "AI based Reading System for Blind using OCR", ISBN: 978-1-7281-0167-5, IEEE Xplore.

[6] M. Geetha and U. C. Manjusha, , A Vision Based Recognition of Indian Sign Language Alphabets and Numerals Using B-Spline Approximation, International Journal on Computer Science and Engineering (IJCSE), vol. 4, no. 3, pp. 406-415. 2012.

[7] Huang, J., Zhou, W., & Li, H. (2015). Sign Language Recognition using 3D convolutional neural networks. IEEE International Conference on Multimedia and Expo (ICME) (pp. 1-6). Turin: IEEE.