IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 3, June 2023

Sign Language to Text Language Conversion

Adnan Md Ashpak Qureshi[1], Athrava Manoj Kargirwar[2], Ishaan Iqbal Sheikh[3], Kartik Sadashiv Tummawar[4], Mohit Sharad Mandhalkar[5], Prof. Anand D. G. Donald[6]

Students, Department of Computer Science & Engineering[1,2,3,4,5]

Guide, Department of Computer Science and Engineering[6]

Rajiv Gandhi College of Engineering, Research & Technology, Chandrapur, Maharashtra, India.

Abstract: *Sign language is a form of communication that uses gestures and gestures to convey meaning. We propose a new method to convert signed words into text. Our system is designed to enable deaf people to communicate with others in a simpler and more convenient way. The plan uses computer vision and deep learning to recognize gestures and translate them into appropriate text. The system was developed using MediaPipe for feature detection, data prioritization, logging, feature generation, and LSTM neural network training. This project has the potential to improve communication skills for the deaf and reduce communication problems with the rest of the world. The system uses key search algorithms such as MediaPipe to recognize traffic and convert it to the corresponding text using the Lstm format. The data collected from the language is preprocessed and then used to train an LSTM neural network to recognize gestures and generate text. This transformation not only helps the deaf and hard of hearing communicate with locals, but is also an aid to those trying to learn the language. Overall, the solution has the potential to improve communication and reduce problems for the deaf and hard of hearing.*

Keywords: Sign Language Character Recognition, Recurrent Neural Network, Computer Vision, Deep Learning

I. INTRODUCTION

The movement of a part of the body (e.g. face, hand) is a gesture. here we use image processing and computer vision to determine point recognized gesture enables computers to understand human behaviour and also acts as an interpreter between computers and humans. this could provide people with the ability to interact with the computer without making physical contact with any equipment. signs are made in the language of congregation for the deaf the community uses sign language in situations where voice is not possible in communication or where it is difficult to type and write but is visible. words were the only means of communication between people at that time. most languages are used when people don't want to speak, but this is the only way of communication in the deaf community. symbols also give the same meaning as spoken words. this is used by the deaf worldwide, but isl, asl, etc. used in regional forms the sign language can be moved with one or both of the hands. it has two separate sign languages and extended instructions. isolated sign language has a word with a gesture while continuous isl or continuous sign language is a series of movements, as the main sentence. in this report, we use standard representational asl hand gestures.

Sign Language

Deaf people around the world communicate using sign language as distinct from spoken language in their everyday a visual language that uses a system of manual, facial and body movements as the means of communication. Sign language is not an universal language, and different sign languages are used in different countries, like the many spoken languages all over the world. Some countries such as Belgium, the UK, the USA or India may have more than one sign language. Hundreds of sign languages are in used around the world, for instance, Japanese Sign Language, British Sign Language (BSL), Spanish Sign Language, Turkish Sign Language.

DOI: 10.48175/IJARSCT-11429

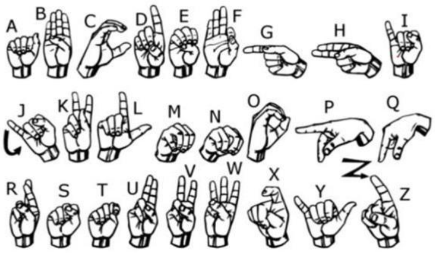


Figure 1: Sign language Alphabets

Recurrent Neural Network (RNN)

Humans don't start their thinking from scratch every second. We don't throw everything away and start thinking from scratch again. Our thoughts have persistence. Traditional neural networks can't do this but Recurrent Neural Networks can. There is information in the sequence itself, and recurrent nets use it to perform tasks that feed forward networks can't. Recurrent networks are distinguished from feed forward networks by the fact that they have feedback loop, ingesting their own outputs moment after moment as input. They're especially useful with sequential data because each neuron or unit can use its internal memory to maintain information about the previous input. For example in case of a network that is suppose to classify what kind of event is happening at every point in a movie. It requires the network to use its reasoning about previous events in the film to inform later ones. Another example in case of language, "I had washed my house" is much more different than "I had my house washed". This allows the network to gain a deeper understanding of the statement. This is important to note because reading through a sentence even as a human, you're picking up the context of each word from the words before it.

Long Short Term Memory Units (LSTMs)

A variation of recurrent net with "Long Short Term Memory Units "LSTMs, was proposed by the German researchers Sepp Hochreiter and Juergen Schmidhuber as a solution to the vanishing gradient problem. LSTMs help preserve the error that can be back propagated through time and layers. By maintaining a more constant error, they allow recurrent nets to continue to learn over many time steps (over 1000), thereby opening a channel to link causes and effects remotely. LSTMs are explicitly designed to avoid the long-term dependency problem. Remembering information for long periods of time is practically their default behavior, not something they struggle to learn!

II. LITERATURE SURVEY

In the recent years, there has been tremendous research on the hand sign language gesture recognition. The technology for gesture recognition is given

Vision based

In the Visual method, the computer camera is an input device that displays data from the hand or finger. The proposed approach requires only one camera enabling human-computer interaction without the use of additional hardware. This system tends to complement Food Vision by identifying vision systems that use in software and/or hardware. This poses a difficult problem because for these systems to be successful, the must have a contrast-free background, an unobstructed view, and human and stand-alone cameras. Additionally, the process must be optimized to meet requirements, including accuracy and robustness. Vision-based analysis is based on how people view information about their environment, but it is the most difficult to use to satisfy Many different methods have been tried so far. One is to create 3D model of the human hand the model maps hand images from one or more cameras and estimates parameters for palm orientation and two angles. is then not used as a deployment guide Second, it uses the camera to capture the image and then extracts some features that are used as input for the classification algorithm that classifies methodology

Data Collection

To generate language-to-text conversion, a large and diverse database of sign language is needed, which is Indian Sign Language. This information was collected with the help of the web and library information line. The Media Pipe library provides tools to track movements in real time and view key details of the user's hand. The webcam captures the movements and stores them as sample data in the dataset.

The data collected by is used to train and test machine learning models responsible for recognizing gestures and converting them to text. To ensure the robustness and accuracy of the system, it is important to collect different data and representations that include various gestures and hand changes. The data collection process is ongoing and the data is constantly updated to ensure that it reflects the Indian Language. With the help of Media Pipe's library and webcams, we can gather the best data to create a powerful and accurate session-to-text conversion.

Data Pre-Processing

Preprocessing motion images is an important step in creating sign-to-text converters. The purpose of preprocessing images is to prepare them for machine learning models, making it easier for them to recognize gestures and translate them into text. In a preliminary step, motion images have been resized, normalized and made suitable for introduction to machine learning models. Resize the images to a smaller size so that the model can make them easier. The normalization step is performed to eliminate inconsistencies in the color of the lighting, background, or image that could adversely affect the model's performance. In addition to resizing and normalizing, the image can also undergo transformation operations such as cropping or rotating to ensure the model has the same orientation. This helps reduce the variance of the data and makes it easier for the model to recognize the direction.

After the preliminary steps are completed, the images can be used to train and test machine learning models. Pre-processed images provide the model with the information it needs to learn the relationship between gestures and text, allowing it to know gestures and translate people to text. With preprocessing, sign-to-text systems are powerful tools that help deaf people communicate with others.

Labeling Text Data

Drawing gestures is an important step in creating a sentence-to-text conversion. In this step, each marker in the data is given a label representing the word or phrase it represents. This registration process is important because it provides the machine learning model with the information it needs to recognize gestures and translate them into text. Motion tags are based on Indian Sign Language and are created according to the standard terminology and grammar used in the language. Tags are submitted by an expert Indian Translator who ensures the tags are consistent and accurate. Although the recording process is done manually, in some cases it can also be done with the help of computer vision techniques. Once movements are recorded, they can be used to train and test machine learning models. Descriptive information provides the model with the information it needs to learn the relationship between gestures and text relationships, enabling it to recognize gestures and interpret letters correctly. With proper recording, session-to-text transcription can be a powerful tool to help deaf people communicate with others.

Training and Testing

During training, the model is fed pre-processed motion images along with written text. The model uses this data to learn the relationship between gestures and text, and updates its parameters a sit processes more data. The purpose of the training phase is to train the model to correctly recognize and transcribe hand gestures**.** In our "Meaning to Text" project, we used a multi-layered LSTM (long-term memory) model to efficiently convert hand gestures into text. The LSTM model consists of three LSTM layers and three Dense layers with RELU activation function and an output layer with SoftMax activation function, all of which help to understand and interpret the nature of the signal. Adding more LSTM layers allows the model to learn and capture patterns and progressions specific to hand gestures. Each LSTM layer in the network consumes a lot of resources, runs from its own memory cells and broadcasts a hidden state that sends the information to the next layer.

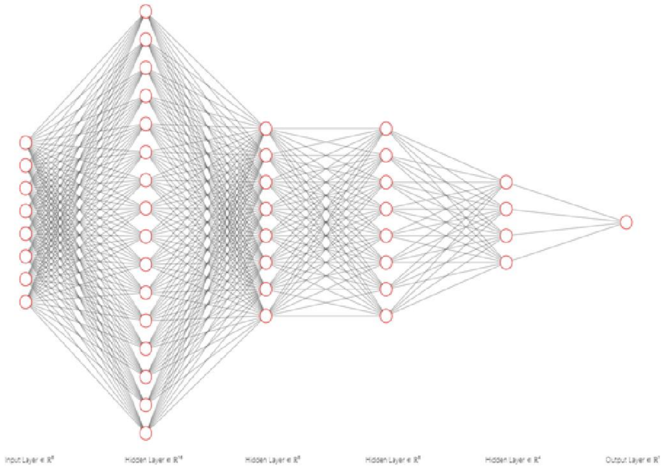


Fig. 2 layers used representation

IMAGE AUGMENTATION AND RESIZE

The image dataset consists of ASL gestures from. The dataset consists of 2080(26*80) images with 80 images per category. Each category represented a different character of English Alphabet. This dataset was then augmented to create a dataset of 1781 images. Out of this dataset 75% i.e. 1500 images were used for training and remaining 25% i.e. 580 images were used for testing. The images in the data set were of a varying size and shape. Therefore the first step was to read and resize each of the images to the similar size of 224x224 pixels. Only when all of the images in the dataset are of the same size can the images be fed into a neural network for training

IMAGE PREPROCESSING

The mean value of RGB over all pixels was subtracted from each pixel value. i.e. in the first pass the model will compute the mean pixel value of each channel over the entire set of pixels in a channel and in the second pass it will modify the images by subtracting the mean from each pixel value. Subtracting the mean value from the pixels centre the data. The mean is subtracted because the model involves. Multiplying weights and adding biases to the initial inputs to cause activations then back propagated with the gradients to train the model.For Image Pre-processing I used the Direct Python function to convert the colour image to gray and remove blur which was used in this project and saving gray images in another directory.

```
frame=cv2.imread(path)
gray=cv2.cvtColor(frame,cv2.COLOR_BGR2GRAY
)
blur = cv2.GaussianBlur(gray,(5,5),2)
```

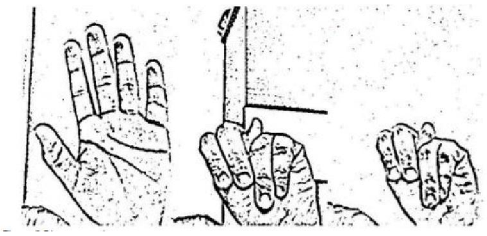


Figure 2: Converting color image to gray scale image

DOI: 10.48175/IJARSCT-11429

Architecture of the Project

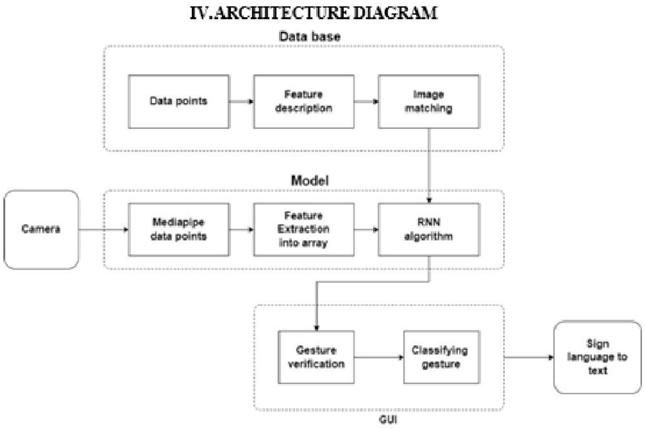
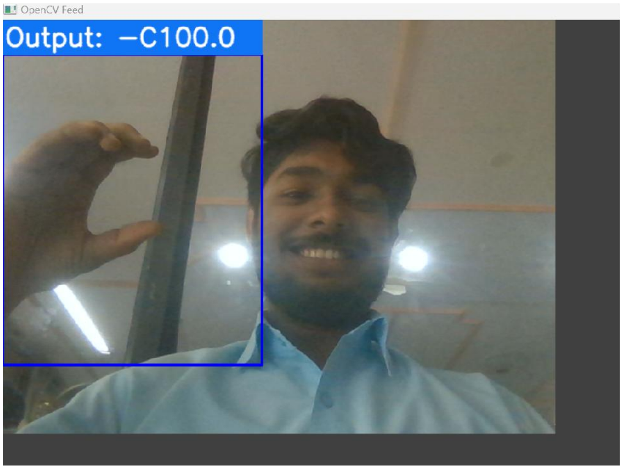


Fig. 3 Architecture Diagram

The conversion of sign language to text involves several steps and technologies, including the use of camera, Mediapipe library, feature extraction, data points, image matching, RNN algorithm, gesture verification, and gesture classification.

III. RESULT

The final output after using rnn, lstm model to convert sign language to text language looks like in the fig 4



Output: –C100.0

Figure 4: final output of the project

IV. CONCLUSION

This project focuses on solving the problems of the deaf. The system will perform the enviable task of recognizing sign language that normal people cannot understand, reducing workload and making working time efficient and accurate. We are trying to build this system using many concepts and libraries for image processing and image features. This article presents an image-based system capable of hand gestures and transcription from sign language. The proposed method has been tested in a real-time situation and it has been seen that the obtained RNN model can recognize the direction.

ACKNOWLEDGEMENT

We would like to thank prof. Anand D.G. Donald sir for their support and guidance in completing our project and helping us in writing and publishing this paper. We would also like to thank our department for supporting us by providing us with the required system.

REFERENCES

[1] S. M Mahesh Kumar, “Conversion od Sign Language into Text,” International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 9, 2018.

[2] Kohsheen Tiku, Jayshree Maloo, Aishwarya Ramesh, Indra R, “Real-time Conversion of Sign Language to Text and Speech,” 2020 Second International Conferenceon Inventive Research in Computing Applications, Coimbatore, India, 2020, pp. 346-351.

[3] C. Uma Bharti, G. Ragavi, K. Karthika \"Signtalk: Sign Language to Text and Speech Conversion,\" 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Coimbatore, India, 2021, pp. 1-4, doi: 10.1109/ICAECA52838.2021.9675751.

[4] M. Zamani and H. R. Kanan, \"Saliency based alphabet and numbers of American sign language recognition using linear feature extraction,\" 2014 4th International Conference on Computer and Knowledge Engineering (ICCKE), Mashhad, Iran, 2014, pp. 398-403, doi: 10.1109/ICCKE.2014.6993442.

[5] A. Saxena, D. K. Jain and A. Singhal, \"Sign Language Recognition Using Principal Component Analysis,\" 2014 Fourth International Conference on Communication Systems and Network Technologies, Bhopal, India, 2014, pp. 810-813, doi: 10.1109/CSNT.2014.168. [6] J. Zhang, W. Zhou, C. Xie, J. Pu and H. Li, \"Chinese sign language recognition with adaptive HMM,\" 2016 IEEE International Conference on Multimedia and Expo (ICME), Seattle, WA, USA, 2016, pp. 1-6, doi: 10.1109/ICME.2016.7552950.

[7] D. Guo, W. Zhou, M. Wang and H. Li, \"Sign language recognition based on adaptive HMMS with data augmentation,\" 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 2016, pp. 2876-2880, doi: 10.1109/ICIP.2016.7532885.

[8] K. Grobel and M. Assan, \"Isolated sign language recognition using hidden Markov models,\" 1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation, Orlando, FL, USA, 1997, pp. 162-167 vol.1, doi: 10.1109/ICSMC.1997.625742.

[9] D. Guo, W. Zhou, M. Wang and H. Li, \"Sign language recognition based on adaptive HMMS with data augmentation,\" 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 2016, pp. 2876-2880, doi: 10.1109/ICIP.2016.7532885.

[10] D. Van Hieu and S. Nitsuwat, \"Image Preprocessing and Trajectory Feature Extraction based on Hidden Markov Models for Sign Language Recognition,\" 2008 Ninth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, Phuket, Thailand, 2008, pp. 501-506, doi: 10.1109/SNPD.2008.80

DOI: 10.48175/IJARSCT-11429