Batch No: 19PB22

REAL TIME SIGN LANGUAGE RECOGNITION USING TRANSFER LEARNING

G Harsha Vardhan
19P61A0566
ghvy999@gmail.com

G Harsha Vardhan rao
19P61A0575
hv5744577@gmail.com

S Harshith
19P61A0581
suryawanshiharshith@gmail.com

VIGNANA BHARATHI INSTITUTE OF TECHNOLOGY
(A UGC Autonomous Institution, Approved by AICTE, Accredited by NBA & NAAC, Affiliated to JNTUH)
Aushapur(V), Ghatkesar(M), Medchal(Dt.), Telangana - 501301

OBJECTIVE:

The main Objective of Sign Language Recognition (SLR) systems is to provide an efficient and accurate way to convert sign language into text or voice has aids for the hearing impaired for example, enabling very young children to interact with computers (recognizing sign language), or among others.

ABSTRACT:

Communication is very crucial to human beings, as it enables us to express ourselves. We communicate through speech, gestures, body language, reading, writing or through visual aids, speech being one of the most commonly used among them. However, unfortunately, for the speaking and hearing-impaired minority, there is a communication gap.Sign language is learned by deaf and mute and usually it is not known to normal people. It strikes our mind that to bridge the gap between hearing impaired and normal people & make communication easier.

To overcome this barrier, we propose a method where we collect sign language gestures using webcam and by using Transfer learning, we train a Tensor Flow model to create a Real-time Sign Language Recognition system. This will help a lot of people in communicating with deaf and mute people.

Keywords : Sign Language Recognition, Transfer Learning, Computer Vision, Deep Learning, Python.

INTRODUCTION:

Sign languages are one of the means of communication through body movements especially of the hands and arms used when spoken type of communication is not possible. It has become the core form of communication for the communities of deaf and mute people. Hence to overcome this barrier of communication among spoken people and sign language users. We have proposed our model which uses a popular deep learning technique called "Transfer Learning" .

Dumb and deaf persons experience difficulties connecting with computers in the workplace because they cannot hear them. It is also risky to travel places alone since they cannot hear cars, bikes, or other people approaching. They can't immediately adapt to their surroundings or respond to other people, and expressing oneself is difficult. Sign language has a long history in western societies as a visual language or technique of communication, dating back to the 17th century. Traditional gestures, mimics, hand signs, and figure spelling, as well as the use of hand position to represent letters of the alphabet, make up sign language. A sign can also represent an entire thought or statement. The major goal is to deliver speech and text output for deaf persons utilising hand gesture sign language without the use of any sensors in a smart method.

TOOLS:

SOFTWARE REQUIREMENTS:

- Language - Python
- Operating System - Windows 8 or above

HARDWARE REQUIREMENTS:

- Processor - i3 or above
- RAM - 8GB
- Hard Disk - 320GB

EXISTING SYSTEM:

Literature review of our proposed system shows that there have been many explorations done to tackle the sign recognition in videos and images using several methods and algorithms. Siming He[4] proposed a system having a dataset of 40 common words and 10,000 sign language images. To locate the hand regions in the video frame, Faster R-CNN with an embedded RPN module is used. It improves performance in terms of accuracy. Detection and template classification can be done at a higher speed as compared to single stage target detection algorithm such as YOLO. The detection accuracy of Faster

R-CNN in the paper increases from 89.0% to 91.7% as compared to Fast-RCNN. A 3D CNN is used for feature extraction and a sign-language recognition framework consisting of long and short time memory (LSTM) coding and decoding network are built for the language image sequences.On the problem of RGB sign language image or video recognition in practical problems , the paper merges the hand locating network, 3D CNN feature extraction network and LSTM encoding and decoding to construct the algorithm for extraction. This paper has achieved a recognition of 99% in common vo-cabulary dataset. Let's approach the research done by Rekha, J[5]. which made use of YCbCr skin model to detect and fragment the skin region of the hand gestures. Using Principal Curvature based Region Detector, the image features are extracted and classified with Multi class SVM, DTW and non-linear KNN. A dataset of 23 Indian Sign Language static alphabet signs were used for training and 25 videos for testing. The experimental result obtained were 94.4%for static and 86.4% for dynamic.

In [6] , a low cost approach has been used for image processing . The capture of images was done with a green background so that during processing, the green colour can be easily subtracted from the RGB colourspace and the image gets converted to black and white. The sign gestures were in Sinhala language. The method that thy have proposed in the study is to map the signs using centroid method. It can map the input gesture with a database irrespective of the hands size and position. The prototype has correctly recognised 92% of the sign gestures. The paper by M. Geetha and U. C. Manjusha[7], make use of 50 specimens of every alphabets and digits in a vision based recognition of Indian Sign Language characters and numerals using B-Spine approximations. The region of interest of the sign gesture is analysed and the boundary is removed . The boundary obtained is further transformed to a B-spline curve by using the MaximumCurvature Points(MCPs) as the Control points. The B-spline curve undergoes a series of smoothening process so fea- tures can be extracted. Support vector machine is used to classify the images and the accuracy is 90.00%. In [8], Pigou used CLAP14 as his dataset [9]. It consists of 20 Italian sign gestures.After preprocessing the images , he used a Convolutional Neural network model having 6 layers for training. It is to be noted that his model

is not a 3D CNN and all the kernels are in 2D. He has used Rectified linear Units (ReLU) as activation functions. Feature extraction is performed by the CNN while classification uses ANN or fully connected layer. His work has achieved an accuracy of 91.70% with an error rate of 8.30%. A similar work was done by J Huang [10].He created his own dataset using Kinect and got a total a total of 25 vocabularies which are used in everyday lives. He then applied a 3D CNN in which all kernels are also in 3D. The input of his model consisted of 5 important channels which are colour-r, colour-b, colour-g, depth and body skeleton. He got an average accuracy of 94.2%. Another research paper on Action recognition topic by the author J.Carriera [11] shares some similarities to sign gesture recognition .He used a transfer learning method for his research As his pre-trained dataset, he used both ImageNet[12] and Kinetic Dataset [9] . After training the pertained models using another two datasets namely UCF-101 [13] and HMDB-51 [14], he then merged the RGB model, flow model, pre-trained Kinetic and pre-trained ImageNet.The accuracy he got on UCF-101 dataset is 98.0% and on HMDB-51 is 80.9%

The Disadvantages are

- Highly Expensive.
- Even if performed by any other ML techniques and With online, they Require tons of images to train for a single gesture .

PROPOSED SYSTEM:

The proposed system can recognize the static word sign and represent its label for better communication. The deaf person should submit a gesture or sign image to the system in the proposed system. The system uses a mat lab image processing technique to analyse the sign input and classifies it for recognised identification. When the input image matches the specifi ed dataset, it then starts the voice media through the system. In addition, the output will be displayed in text format. This is a working prototype for the conversion of sign language to speech and text.

Hence , We are implementing using a popular Deep Learning technique known as “Transfer Learning” which help us fasten the training process and also require very less number of images for training the model .

CONCLUSION:

The main purpose of sign language detection system is providing a feasible way of communication between a normal and dumb people by using hand gesture. The proposed system can be accessed by using webcam or any in-built camera that detects the signs and processes them for recognition.

From the result of the model, we can conclude that the proposed system can give accurate results under controlled light and intensity. Furthermore, custom gestures can easily be added and morethe images taken at different angle and frame will provide more accuracy to the model. Thus, the model can easily be extended on a large scale by increasing the dataset.

The model has some limitation such as environmental factors like low light intensity and uncontrolled background which cause decrease in the accuracy of the detection. Therefore, we'll work next to overcome these flaws and also increase the dataset for more accurate results.

REFERENCES:

1. K. Bantupalli and Y. Xie, "American Sign Language Recognition using Deep Learning and Computer Vision," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 4896-4899, doi: 10.1109/BigData.2018.8622141.

2. CABRERA, MARIA & BOGADO, JUAN & FermÃn, Leonardo & AcuÃ±a, Raul & RALEV, DIMITAR. (2012). GLOVE-BASED GESTURE RECOGNI-TIONSYSTEM.

3. He, Siming. (2019). Research of a Sign Language Translation System Based on Deep Learning. 392-396. 10.1109/AIAM48774.2019.00083. International Conference on Trendz in Information Sciences and Computing (TISC). : 30-35, 2012.

4. Herath, H.C.M. & W.A.L.V.Kumari, & Senevirathne, W.A.P.B & Dissanayake, Maheshi. (2013). IMAGE BASED SIGN LANGUAGE RECOGNITION SYSTEM FOR SINHALA SIGN LANGUAGE

5. M. Geetha and U. C. Manjusha, , A Vision Based Recognition of Indian Sign Language Alphabets and Numerals Using B-Spline Approximation, Inter- national Journal on Computer Science and

Engineering (IJCSE), vol. 4, no. 3, pp. 406-415. 2012.

6. Pigou L., Dieleman S., Kindermans PJ., Schrauwen B. (2015) Sign Language Recognition Using Convolutional Neural Networks. In: Agapito L., Bronstein M., Rother C. (eds) Computer Vision – ECCV 2014 Workshops. ECCV 2014. Lecture Notes in Computer Science, vol 8925. Springer, Cham. https://doi.org/10.1007/978-3-319-16178-5_40

7. Escalera, S., BarÃ³, X., GonzÃ lez, J., Bautista, M., Madadi, M., Reyes, M., . . . Guyon, I. (2014). ChaLearn Looking at People Challenge 2014: Dataset and Results. Workshop at the European Conference on Computer Vision (pp. 459-473). Springer, . Cham.

8. 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.

9. Jaoa Carriera, A. Z. (2018). Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on (pp. 4724-4733). IEEE. Honolulu.

Date of Verification:

Date of Submission:

Guide Signature: