International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Survey on Sign language to text and speech conversion using Machine Learning model

Sujay R[a], *Somashekar M*[b], *Aruna Rao B P*[c]

[a]*BE Student, ECE Dept., K S Institute of Technology, No.14, Raghuvanahalli, Kanakapura Main Road, Bengaluru, Karnataka 560109, India*
[b]*BE Student, ECE Dept., K S Institute of Technology, No.14, Raghuvanahalli, Kanakapura Main Road, Bengaluru, Karnataka 560109, India*
[c]*Assitant Professor, ECE Dept, K S Institute of Technology, No.14, Raghuvanahalli, Kanakapura Main Road, Bengaluru, Karnataka 560109, India*

ABSTRACT

Communication provides interaction among people to exchange feelings and ideas. The deaf community suffers a lot from interacting with the community. Sign language is a way through which people communicate with the deaf. In order to provide interaction with people who are unaware of the language, there is a system that can convert the sign languages to understandable form. The purpose of this work is to provide a real-time system that can convert Sign Language to text. In this paper, we are introducing a deep learning approach which can classify the sign using the convolutional neural network. In the first phase, we make a classifier model using the numeral signs using the Keras implementation of a convolutional neural network using python. In phase two another real-time system which is used to convert Alphabet signs to text using the same process. A frame is extracted from a continuous video feed and it is fed to the trained CNN model to provide the corresponding and desired output.

Keywords:Machine Learning, Convolutional Neural Network, Sign Language, Literature Survey.

1. Introduction

Sign language is a basic means of communication for those with hearing and vocal disabilities. Those disadvantaged face difficulty in their day to day lives to communicate with others. We are aiming to develop a system that would eradicate this barrier in communication. Sign language consists of making shapes or movements with our hands with respect to one's head or other body parts along with certain facial cues. A recognition system would thus have to identify the head and hand orientation or the hand movements, facial expression and even body pose of the Signer.

An effective hand-based interaction method that involves recognition of hand gestures using Machine Learning to predict letters in the form of text \ video conveying what the user is trying to express. The steps include Pre-processing of images extracted from the video feed, converting the user's hand and eliminating the background and comparing the hand image with the dataset using the model leading to gesture recognition. The predicted text can be converted to speech or displayed on a monitor. We will be using OpenCV to capture and pre-process images required to create the dataset, TensorFlow to train and predict using CNN (Convolutional Neural Network) model.

* *Somashekar M.*
E-mail address: shekarmeda342@gmail.com

2. Implementation

2.1. Data Acquisition

In sign language recognition we have to classify the input gestures into respective alphabets/numerical. A Classifier algorithm is used, it falls under supervised machine learning. Where the data is pre-categorized, Labelled Training and testing data have to be given as input to train the model. Data for training can be obtained from various methods and sources.

Datasets can be obtained from various sources on the internet. In [1] they have used the 2012 ILSVRC dataset to train the model. ILSVRC is composed of 10000 uniquely different objects and classes. Many datasets of high-quality images and videos ASL in available in Kaggle e.g., MINST Sign language.

Dataset can be created by considering certain conditions such as lighting, angle, skin colour etc. In [2] selfie sign language dataset is created using a selfie camera. The dataset is having 200 ISL at different viewing angles and at 30fps. Creating our own data set depends on the resolution of the camera .and the captured images have to be pre-processed to remove the noise in the images

A 3rd dimension depth can be added. In [3] Kinect is used to capture the images; it is a motion sensor that can provide colour stream and depth. This depth feature helps in identifying the background and foreground objects, which increases the accuracy of classification.

If the dataset acquired is small data augmentation [1]can be done. In data augmentation of the datasets resulted in performance by 20% increase in accuracy, the dataset was increased by slight changes such few pixels changes, flipping of the images

2.2. Pre-Processing

Pre-processing of data improves the image quality and remove the distortion and noise in the images. It is the first step after the data acquisition.

Resizing of the images is done in almost every approach[1]–[6] it helps to give consistent input to the model

In [2] padding the images with black pixels was done, this helped to preserve the aspect ration after resizing and remove fever relevant pixels from random crops. Background subtraction techniques[1] are used to remove the background of the image.

CLACHE (Contrast Limited Adaptive Histogram Equalization) [6] was applied to the dataset images to equalize the lightness in the image frame using LAB colour systema and reduce the noise amplification.

Gaussian blurring[6] is used to apply blurring on theimage. To obtain the skin, thresholding operation using the HSV colour space is applied to images

2.3. Algorithm

CNN is the most popular choice for classification process in sign to language recognition because of its high accuracy.

Transfer Learning: Transfer Learning[2] is a machine learning technique where models are trained on (usually) larger data sets and refactored to fit more specific or niche dataset. In this method. Caffe and GoogLeNet deep CNN developed by Google pre-trained 2012 ILSVRC dataset is used to train their own data set. Pre trained will lead in increase in accuracy and less training time.

3 convolutional layers and 3 pooling layers is used in activation function called RELU is used in [6]between these convolutional and pooling layers. In one of the architecture four [5]convolutional layers, five rectified linear units (ReLu), two stochastic pooling layers, one dense and one SoftMax output layer was used and stochastic pooling was considered as the best pooling as it showed increase in accuracy compared to other pooling techniques.

3D Convolution neural network[4]: Here in five types of input data is given with RGB three channels, depth and body skeleton which results in 5 feature maps denoted by colour-R, colour-G, colour-B, depth, body skeleton. A 3D kernel is used on the convolution layers and the architecture also contains subsampling after each convolution layer

In one of the methods [3]AdaBoost a machine learning algorithm is used for feature extraction. To classify the extracted features Haar-Cascade algorithm is used. The features are compared with the stored features of every sign’s feature in the trained database (trained using Haar-Cascade). Haar-like features based AdaBoost classifier showed robustness and fastness in different lightning, scaling change as well as rotation.

2.4. Output

Text of the recognized gestures is displayed on the desktop screen in all. of the projects reviewed.

SAPI 5.3 was used in [3] to covert text to speech, SAPI is a speech API created by Microsoft for voice recognition and synthesis. Two more signs were trained, SPACE sign which helps to separate words and OK sign which is used for the sign of finishing capture the video frames and start to convert the text to speech.

2.5. Comparison Table

Reference	Pre-processing	Model	Data set	Accuracy*
5	Resizing images	4 CNN layers 5 linear units 2 Pooling layers 1 Dense and Softmax	Selfie Sign Language Database	92.88
6	CLACHE Gaussian Blur Thresholding using HSV	CNN ReLu	ISL	99.56
2	Padding of images using Black pixels	Caffe and GoogleNet 2 / 3 layers using Xavier initialization	ASL 2012 ILSRV dataset	
3	Algorithm to find the Feature vector Noise reduction, Gray scaling	Haar Cascade algorithm	ASL	98.7
1	Background subtraction Data Augmentation	3 groups of 2 CNN followed by max pool layer, dropout followed by a connected layer, a dropout layer	Data augmentation and own dataset	82.5
4	Feature map denoted by RGB Depth Body skeleton	CNN model GMM-HMM model	Own dataset	94.2

**Accuracy in %*

3. Conclusion

In this paper, we have compared the various techniques used for real-time Sign Language recognition using Machine Learning. We have compared the Pre-processing techniques of the input video feed, methods used to train the Machine Learning model and the accuracy of each method. We see that the Algorithms used with Neural Network have higher accuracy only when the training data-set is large and the Algorithm works efficiently when it is trained by a larger data-set as compared to a smaller data-set. The main challenge faced in Sign language recognition is the detection of fingerspells. Various techniques have been employed for hand segmentation, background removal and to pre-process the images. The other factors affecting the recognition of fingerspells are the colour of the Background, angle of the wrist, Quality of the camera used. This paper provides the various techniques which can be employed for real-time Sign language recognition along with their shortcomings and guidelines to the new researchers.

References

[1] L. State, "UsingDeep ConvolutionalNetworksfor Gesture Recognition in American Sign Language".

[2] B. Garcia, "Real-time American Sign Language Recognition with Convolutional Neural Networks Sigberto Alarcon Viesca Stanford University," 2016.

[3] V. N. T. Truong, C. K. Yang, and Q. V. Tran, "A translator for American sign language to text and speech",*2016 IEEE 5th Global Conference on Consumer Electronics, GCCE 2016*, DOI: 10.1109/GCCE.2016.7800427

[4] "SIGN LANGUAGE RECOGNITION USING 3D CONVOLUTIONAL NEURAL NETWORKS' Jie Huang, Wengang Zhou, Houqiang Li, and Weiping Li University of Science and Technology of China, Hefei, China.

[5] G. A. Rao, K. Syamala, P. V. V. Kishore, and A. S. C. S. Sastry, "Deep convolutional neural networks for sign language recognition",*2018 Conferenceon Signal Processing andCommunincation Engineering Systems, SPACES 2018*, vol. 2018-January, pp. 194–197, 2018, DOI: 10.1109/SPACES.2018.8316344.

[6] T. D. Sajanraj and M. V. Beena, "Indian Sign Language Numeral Recognition Using Region of Interest Convolutional Neural Network",*Proc. Int. Conf. Inven. Commun. Comput. Technol. ICICCT 2018*, no. Icicct, pp. 636–640, 2018, DOI: 10.1109/ICICCT.2018.8473141.