Sign Language to Text Conversion using CNN (For Deaf and Mute People)

Abhay Kumar, Anurag Gupta ,Bipin Chaurasia ,Chaitanya Mishra

Under the guidance of - Apoorv Mishra sir
Computer Science & Engineering Department,
Maharana Pratap Engineering College
Kanpur, India

Abstract:

There is an undeniable communication problem between the Deaf community and the hearing majority. Sign language is one of the most natural form of language used by deaf and dumb people to communicate, but since most people do not know sign language and interpreters are very difficult to come by. Hence we have come up with a real time method using various technologies like convolution neural network, artificial neural network for finger spelling based American Sign Language. Various libraries like TensorFlow (Open source software library for numerical computations), OpenCV (Open source library of programming functions used for real time computer vision), Keras (High level neural networks library written in Python that works as a refer to TensorFlow) are used. In our method, the hand is first passed through a filter and after the filter hand is passed through a classifier which predicts the class of hand gestures. Our method provides 95.7% accuracy for the 26 letters of alphabet.

Our project aims to create a computer application and train a model which when shown a real time video of hand gestures of American Sign Language shows the output for that particular sign in text format on the screen.

Sign Language Recognition is one of the most growing fields of research area. Many new techniques have been developed recently in this area. The Sign Language is mainly used for communication of deaf-dumb people. This project shows the sign language recognizing of 26 hand gestures in American sign language using CNN. The proposed system contains four modules such as: pre- processing and hand segmentation, feature extraction, sign recognition and sign to text. By using image processing the segmentation can be done.

Introduction:

American sign language is a predominant sign language Since the only disability D&M people have is communication related and they cannot use spoken languages hence the only way for them to communicate is through sign language. Communication is the process of exchange of thoughts and messages in various ways such as speech, signals, behavior and visuals. Deaf and dumb(D&M) people make use of their hands to express different gestures to express their ideas with other people. Gestures are the nonverbally exchanged messages and these gestures are understood with vision. This nonverbal communication of deaf and dumb people is called sign language.

American Sign Language (ASL) is natural syntax that has the same etymological homes as being speaking languages, having completely different grammar, ASL can be express with destiny of actions of the body. In native America, people who are deaf or cant see, its a reliable source of absurdity. There is not any formal or familiar form of sign language. Different signal languages are speculating in particular areas. For a case, British Sign Language (BSL) is an entirely different language from an ASL, and USA people who familiarise with ASL would not easily understand BSL. Some nations adopt capabilities of ASL of their sign languages. Sign language is a way of verbal exchange via human beings diminished by speech and listening to loss. Around 360 million human beings globally be afflicted via unable to hearing loss out of which 328000000 are adults and 32000000 children. hearing impairment extra than 40 decibels in the better listening to ear is referred as disabling listening to loss.

Literature Review:

In the recent years there has been tremendous research done on the hand gesture recognition. With the help of literature survey done we realized the basic steps in hand gesture recognition are:-

• Data acquisition

• Data preprocessing

• Feature extraction

• Gesture classification

• S. S Kumar, T. Wangyal, V. Saboo and R. Srinath, "Time Series Neural Networks for Real Time Sign Language Translation," 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, 2018, pp. 243-248, doi: 10.1109/ICMLA.2018.00043.

• E. Abraham, A. Nayak and A. Iqbal, "Real-Time Translation of Indian Sign Language using LSTM," 2019 Global Conference for Advancement in Technology (GCAT), BANGALURU, India, 2019, pp. 1-5, doi: 10.1109/GCAT47503.2019.8978343.

It uses electromechanical devices to provide exact hand Configuration and position. Different glove based approaches can be Used to extract information.
But it is expensive and not user friendly.

Vision based approach

- In vision based methods computer camera is the input device for observing the information of hands or fingers. The Vision Based methods require only a camera, thus realizing a natural interaction between humans and computers without the use of any extra devices. These systems tend to complement biological vision by describing artificial vision systems that are implemented in software and/or hardware.
- The gestures are captured through the web camera. This OpenCV video stream is used to capture the entire signing duration. The frames are extracted from the stream and are processed as gray scale images with the dimension of 50*50. This dimension is consistent throughout the project as the entire dataset is sized exactly.
- The model accumulates the recognized gesture to words. The recognized words are converted into the corresponding speech using the pyttsx3 library. The text to speech result is a simple work around but is an invaluable feature as it gives a feel of an actual verbal conversation.

Methodology:

This system is a vision based approach. All the signs are represented with bare hands and so it eliminates the problem of using any artificial devices for interaction.

Data Set Generation

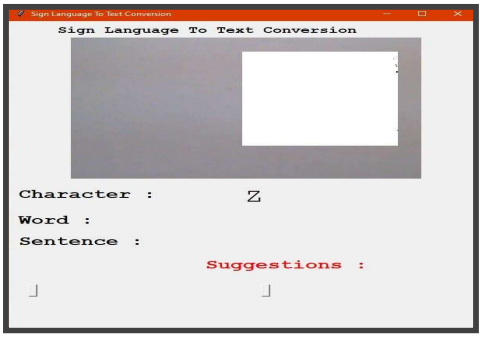
For the project we tried to find already made datasets but we couldn't find dataset in the form of raw images that matched our requirements. All we could find were the datasets in the form of RGB values. Hence we decided to create our own data set. Steps we followed to create our data set are as follows.
We used Open computer vision(OpenCV) library in order to produce our dataset. Firstly we captured around 800 images of each of the symbol in ASL for training purposes and around 200 images per symbol for testing purpose. First we capture each frame shown by the webcam of our machine. In the each frame we define a region of interest (ROI) which is denoted by a blue bounded square as shown in the image below.

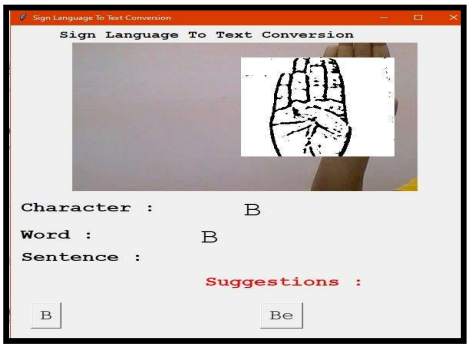
From this whole image we extract our ROI which is RGB and convert it into gray scale Image as shown below.

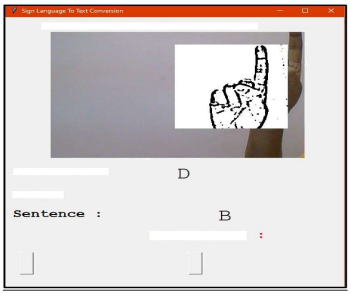
Finally we apply our Gaussian blur filter to our image which helps us extracting various features of our image. The image after applying Gaussian blur looks like below.

Results and discussion:

We have achieved an accuracy of 95.8% in our model using only layer 1 of our algorithm , and using the combination of layer 1 and layer 2 we achieve an accuracy of 98.0%, which is a better accuracy then most of the current research papers on american sign language. Most of the research papers focus on using devices like kinect for hand detection. In [7] they build a recognition system for flemish sign language using convolutional neural networks and kinect and achieve an error rate of 2.5%. In [8] a recognition model is built using hidden markov model classifier and a vocabulary of 30 words and they achieve an error rate of 10.90%. In [9] they achieve an average accuracy of 86% for 41 static gestures in japanese sign language. Using depth sensors map [10] achieved an accuracy of 99.99% for observed signers and 83.58% and 85.49% for new signers. They also used CNN for their recognition system. One thing should be noted that our model doesn't uses any background subtraction algorithm whiles some of the models present above do that. So once we try to implement background subtraction in our project the accuracies may vary. On the other hand most of the above projects use kinect devices but our main aim was to create a project which can be used with readily available resources. A sensor like kinect not only isn't readily available but also is expensive for most of audience to buy and our model uses a normal webcam of the laptop hence it is great plus point.. Below are the confusion matrices for our results.







Limitations and Future Research:

There were many challenges faced by us during the project:

- The very first issue we faced was of dataset. We wanted to deal with raw images and that
- too square images as CNN in Keras as it was a lot more convenient working with only square images. We couldn't find any existing dataset for that hence we decided to make our own dataset.

- Second issue was to select a filter which we could apply on our images so that proper features of the images could be obtained and hence then we could provided that image as
- input for CNN model.

- We tried various filter including binary threshold, canny edge detection, gaussian blur etc. but finally we settled with gaussian blur filter.

- More issues were faced relating to the accuracy of the model we trained in earlier phases which we eventually improved by increasing the input image size and also by improving the dataset.

- We are planning to achieve higher accuracy even in case of complex backgrounds by trying out various background subtraction algorithms . We are also thinking of improving the preprocessing to predict gestures in low light conditions with a higher accuracy.

- We look forward to use more alphabets 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 also like to enhance the system by adding speech recognition so that blind people can benefit as well.

- We look forward to add audio feature which spell each word shown by hand gestures.

Conclusion:

- In this report, a functional real time vision based American sign language recognition for D&M people have been developed for asl alphabets. We achieved final accuracy of 98.0% on our dataset. We are able to improve our prediction after implementing two layers of algorithms in which we verify and predict symbols which are more similar to each other.

- This way we are able to detect almost all the symbols provided that they are shown properly, there is no noise in the background and lighting is adequate.

- The project is a simple demonstration of how CNN can be used to solve computer vision problems with an extremely high degree of accuracy. A finger spelling sign language translator is obtained which has an accuracy of 95%.

- The project can be extended to other sign languages by building the corresponding dataset and training the CNN. Sign languages are spoken more in context rather than as finger spelling languages, thus, the project is able to solve a subset of the Sign Language translation problem.

- The main objective has been achieved, that is, the need for an interpreter has been eliminated. There are a few finer points that need to be considered when we are running the project.

- The thresh needs to be monitored so that we dont get distorted grayscales in the frames. If this issue is encountered, we need to either reset the histogram or look for places with suitable lighting conditions. We could also use gloves to eliminate the problem of varying skin complexion of the signee. In this project, we could achieve accurate prediction once we started testing using a glove.

- The other issue that people might face is regarding their proficiency in knowing the ASL gestures. Bad gesture postures will not yield correct prediction. This project can be enhanced in a few ways in the future, it could be built as a web or a mobile .

References:

- L. Ku, W. Su, P. Yu and S. Wei, "A real-time portable sign language translation system," 2015 IEEE 58th International Midwest Symposium on Circuits and Systems (MWSCAS), Fort Collins, CO, 2015, pp. 1-4, doi: 10.1109/MWSCAS.2015.7282137.

- S. Shahriar et al., "Real-Time American Sign Language Recognition Using Skin Segmentation and Image Category Classification with Convolutional Neural Network and Deep Learning," TENCON 2018 – 2018 IEEE Region 10 Conference, Jeju, Korea (South), 2018, pp. 1168-1171, doi: 10.1109/TENCON.2018.8650524.

- M. S. Nair, A. P. Nimitha and S. M. Idicula, "Conversion of Malayalam text to Indian sign language using synthetic animation," 2016 International Conference on Next Generation Intelligent Systems (ICNGIS), Kottayam, 2016, pp. 1-4, doi: 10.1109/ICNGIS.2016.7854002.

- M. Mahesh, A. Jayaprakash and M. Geetha, "Sign language translator for mobile platforms," 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, 2017, pp. 1176-1181, doi: 10.1109/ICACCI.2017.8126001.

- S. S Kumar, T. Wangyal, V. Saboo and R. Srinath, "Time Series Neural Networks for Real Time Sign Language Translation," 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, 2018, pp. 243-248, doi: 10.1109/ICMLA.2018.00043.

- D. Kelly, J. Mc Donald and C. Markham, "Weakly Supervised Training of a Sign Language Recognition System Using Multiple Instance Learning Density Matrices," in IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 41, no. 2, pp. 526-541, April2011,doi:10.1109/TSMCB.2010.2065802.