**Sign Language Detection and Conversion to Text Using CNN and OpenCV**

**Debayan Das**

**19MCA0070**

**Project Guide: Prof. M. Varalakshmi**

**Abstract:**

Computer vision has been rose as an important domain of research nowadays. With the technological trend in man-machine interfaces and the machine intelligence, these powers are used for making the life of the people easier and less complex. Communication plays an important role in our daily life, but when it comes to the area of deaf and dumb people it becomes very difficult. The only way out from this is sign language which is uninterpretable by people. This paper focuses on making an interface between the normal people and that deaf and dumb people with the help of computer vision. To do so first a classification model is built to identify the sign alphabet using CNN and then for real time application the hand gesture sign is detected using OpenCV and fed to the model to make the final prediction. And after recognition it is converted to text for further use.

**Introduction:**

Sign language is one of the specific areas of human gesture communication and a complexed language that is used by the various deaf and dumb communities around the world. Unfortunately, there are very few people with proper knowledge of sign language and it leads the deaf and dumb community to a social isolation. Motivated by this here tried to develop a system that would be able to interpret the sign language and that would be helpful for the deaf and dumb community to communicate with natural people without knowledge of sign language.

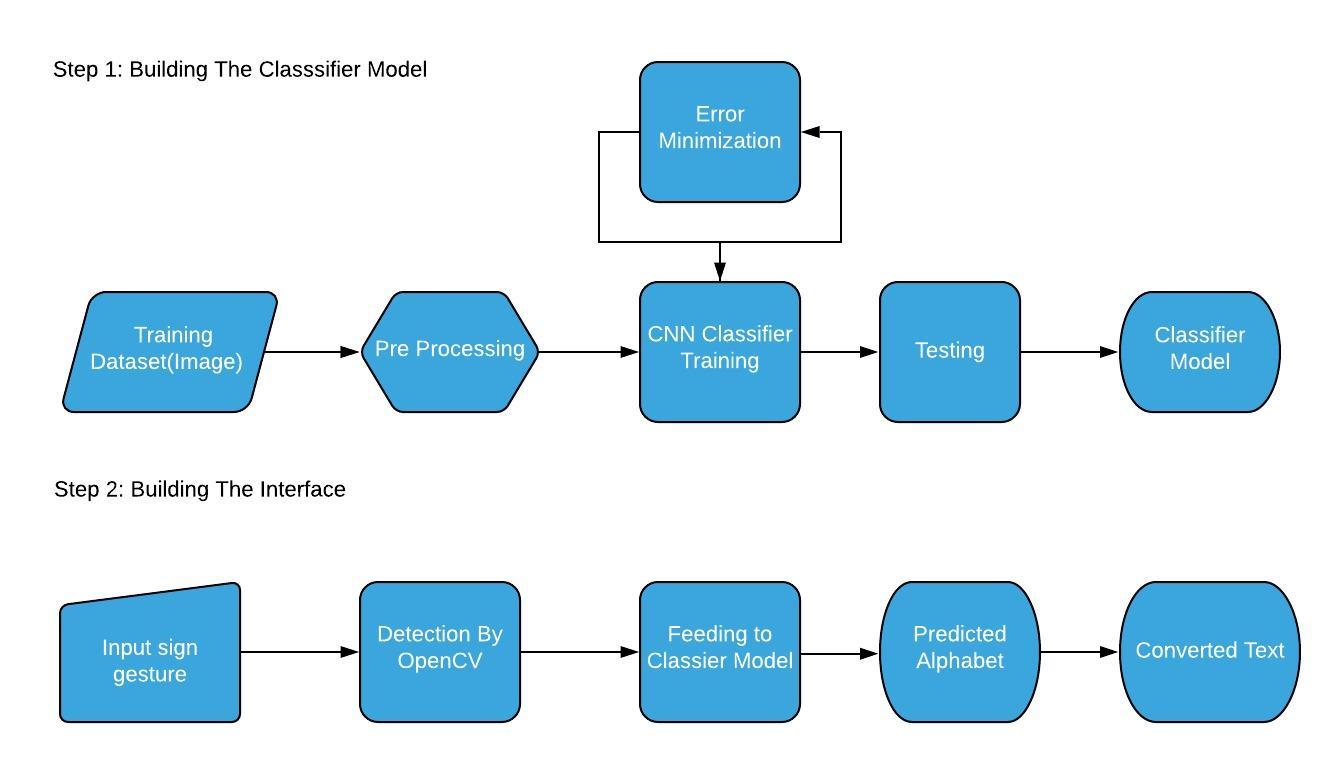
The field of deep learning rapidly growing in recent years and an important domain for research work. Particularly Convolutional Neural Networks are able to achieve a better result in terms of accuracy in the field of Image Classification. Similarly, OpenCV has become a popular open source library in terms of image processing and improving many fields of Image Processing.

In this work a combination of CNN and OpenCV is used to make an interface to make the communication easier between the deaf and dumb community and the people with no proper knowledge of sign language. In this work American Sign Language dataset is used to train the CNN model. Then the interface is made with the help of OpenCV that captures the image and makes it ready so that it can fed to the model. Then the image is fed to the CNN model and the output is displayed in the screen.

**Literature Survey:**

1. In one study [1] an interface is made to make the conversation easier between a normal people and the one with disabilities. For this they have used CNN model. They have trained the model with the hand gesture images and then the have deployed the model with the interface. The detects the gesture from the one with the disabilities and then convert it to text and displays it and converts into speech for both way communication.
2. In another study [2] an interface is created to convert sign language to text using CNN model by data augmentation technique. They images were captured by Microsoft Kinect. The images were augmented to generate more perspective views to avoid the overfitting.
3. In another approach [3] a alignment framework is proposed with iterative optimization. The framework consists of two modules: a 3D-Resnet which is used for feature learning and CTC an encoder decoder sequence learning network where two decoders(LSTM and CTC) are trained together with maximum likelihood criterion. The warping path, which indicates the possible alignment between input video clips and sign words, is used to fine-tune the 3D-ResNet as training labels with classification loss. After fine-tuning, the improved features are extracted for optimization of encoder decoder sequence learning network in next iteration.
4. In this paper [4] they propose RGB and RGB-D static gesture recognition method by the use of a fine-tuned VGG19 Model which uses a feature concatenate layer of RGB and RGB-D images to increase the accuracy. They got 94.8% accuracy on implementing the model on American Sign Language recognition dataset.
5. In this work [5] a framework is proposed to interpret sign language and convert it to text. For this a CNN model is used which is trained by preprocessed hand gesture image dataset.
6. In this study a deep neural network model is used to convert sign videos into natural language sentences by the utilization of human keypoint extraction like face, hand and gesture recognition. For this KETI (Korea Electronics Technology Institute) sign language dataset is used. The translation model achieved 93.28% accuracy.
7. In another study [7] a support vector machine (SVM) based recognition is used to identify hand gestures. The model uses a eigen space size function and hu moments features to classify different hand features.
8. In this study [8] they have used 3D-CNN model to recognize the hand gestures and to convert them to text. They captured the images frames using OpenCV and trained the CNN model with the frames.

**Work Plan:**



**Module Description:**

This work has 3 modules.

1. **Preprocessing:** The first task is to preprocess the images. In order to do so first a background of dark is pasted around all the images and the images are converted to ‘RGB’ format to make the image channels similar. Next all the images are transformed to similar dimension 400x400. At the end images are saved but while saving the images extension of the images are changed to .jpeg from .png.
2. **Training the CNN model:** In the second step the CNN model is trained with the images which are classified to 36 classes (A-Z and 0-9) previously.
3. **Making the Interface:** The 3rd step is to make the interface that will interpret the sign language and will convert to text. This will be made using OpenCV library of python. The image will be captured and will be made ready so that it can be fade to the CNN model and the output of the model will be displayed in the screen.

**Work Progress:**

Till now the first two modules have been completed. The images are preprocessed and the classifier model is trained with the processed images. The model is giving almost 99% accuracy. The CNN model contain total 9 layers. First two are two convolution layers followed by 1 maxpool layer. Then again two convolution layer followed by 1 maxpool layer. Then there is on flatten layer which is followed by two fully connected layer(dense). For optimizing the model stochastic gradient descent(SGD) method is used and the model is trained for 10 epochs.

**References:**

1. Jose, D. A. A. R. D., & Davis, J. (2019). Sign Language Translator Using CNN Model.
2. Tao, W., Leu, M. C., & Yin, Z. (2018). American Sign Language alphabet recognition using Convolutional Neural Networks with multiview augmentation and inference fusion. Engineering Applications of Artificial Intelligence, 76, 202-213.
3. Pu, J., Zhou, W., & Li, H. (2019). Iterative alignment network for continuous sign language recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4165-4174).
4. Khari, M., Garg, A. K., Crespo, R. G., & Verdú, E. (2019). Gesture Recognition of RGB and RGB-D Static Images Using Convolutional Neural Networks. International Journal of Interactive Multimedia & Artificial Intelligence, 5(7).
5. Ghaste, P. V., Bastwade, M. S., Khandelwal, R., Ambapkar, S., Ansari, Z., & Salaria, C. S. (2019). Sign Language Interpretation and Conversion to Text (September 2019). *National Journal of Computer and Applied Science*, *2*(3), 10-13.
6. Ko, S. K., Kim, C. J., Jung, H., & Cho, C. (2019). Neural sign language translation based on human keypoint estimation. *Applied Sciences*, *9*(13), 2683.
7. Kelly, D., McDonald, J., & Markham, C. (2010). A person independent system for recognition of hand postures used in sign language. *Pattern Recognition Letters*, *31*(11), 1359-1368.
8. Ismunandar, A. A. Recognizing Sign Languages Using Pattern Recognition.