Real Time AI Detection on Sign Language Recognition using Convolutional Neural Network with Canny Edge Detection

[1]Shaik Muhammed Shamstabreaz, [2]Muatasim Ahmed, [3]B Venkatesh Prasad, [4]Meripo Anil Babu, [5]Ms.S.Nivetha Lakshmi

[1]Student,[2] Student,[3] Student,[4] Student,[5]Assistant Professor
[1]Department of Computer Science and Engineering,
[1]Bharath Institute of Higher Education and Research, Chennai, India

Abstract - Sign Language is one of the ways to communicate with deaf people. These work-sets (Sign Language), including features and variation in the language with locality have been the major barriers which has led to little research being done in ISL (Indian Sign Language). One should learn sign language to interact with them. Learning usually takes place in peer groups. There are very few study materials available for sign learning. Because of this, the process of learning sign language is a very difficult task. The initial stage of sign learning is Finger spelled sign learning and moreover, is used when no corresponding sign exists, or the signer is not aware of it. We introduce a new technique named canny edge detection (CED), the required results with more accuracy.

Index Terms CNN (Convolution Neural Network), CED (Canny Edge Detection), ISL (Indian Sign Language).

I. INTRODUCTION (HEADING 1)

Communication is one of the basic requirements for survival in society. Deaf and dumb people communicate among themselves using sign language, but normal people find it difficult to understand their language. Extensive work has been done on American sign language recognition but Indian sign language differs significantly from American sign language. Indian Sign Language (ISL) uses two hands for communicating whereas American Sign Language (ASL) uses single hand for communicating. Using both hands often lead to security of features due to overlapping of hands. In addition to this, lack of datasets along with variance in sign language with locality has resulted restrained efforts in ISL gesture detection. Sign language recognition using image-based hand gesture recognition techniques. Hand gesture is one of the methods used in sign language for non-verbal communication. It is most commonly used by deaf & dumb people who have hearing or speech problems to communicate among themselves or with normal people. Sign Language Recognition is a breakthrough for helping deaf-mute people and has been researched for many years. Unfortunately, every research has its own limitations and are still unable to be used commercially.

II. LITERATURE SURVEY

Tanuj Bohra [1] proposed a real-time two-way sign language communication system built using image processing, deep learning and computer vision. Techniques such as hand detection, skin color segmentation, median blur and contour detection are performed on images in the dataset for better results. CNN model trained with a large dataset for 40 classes and was able to predict 17600 test images in 14 seconds with an accuracy of 99%.

Joyeeta Singha and Karen Das [2] proposed a system for Indian sign language recognition from a live video. The system comprises three stages. Preprocessing stage includes skin filtering and histogram matching. Eigenvalues and eigenvectors are being considered for feature extraction stage and Eigen value weighted Euclidean distance for classification. Dataset consisted of 480 images of 24 ISL signs of ISL signed by 20 people. System was tested on 20 videos and achieved an accuracy of 96.25%.

Muthu Mariappan H. and Dr. Gomathi V [3] designed a real time sign language recognition system as a portable unit using contour detection and fuzzy c-means algorithm. Contours are used for detecting face, left and right hand. While a fuzzy c-means algorithm is used to partition the input data into a specified number of clusters. System was implemented on a dataset that contained videos recorded from 10 signers for several words and sentences. It was able to achieve an accuracy of 75%.

Salma Hayani [4] proposed an Arab sign language recognition system based on CNN, inspired from LeNet-5. Dataset contained 7869 images of Arab signs of numbers and letters. Various experiments were conducted by varying the number of training sets from 50% to 80%. 90% accuracy was obtained with 80% training dataset. The author has also compared the results obtained with machine learning algorithms like KNN (k-nearest neighbour) and SVM (support vector machine) to show performance of the system. This model was purely image based and it can be extended to video-based recognition.

Kshitij Bantupalli and Ying Xie [5] worked on an American sign language recognition system which works on video sequences based on CNN, LSTM and RNN. A CNN model named Inception was used to extract spatial features from frames, LSTM for longer time dependencies and RNN to extract temporal features. Various experiments were conducted with varying sample sizes and the dataset consists of 100 different signs performed by 5 signers and maximum accuracy of 93% was obtained. Sequence is then fed to a LSTM for longer time dependencies. Outputs of SoftMax layer and max pooling layer are fed to RNN architecture to extract temporal features from SoftMax layer.

III. PROPOSED SYSTEM

On collected dataset, divided the approach to tackle the classification problem into three stages.

- The first stage is to segment the skin part from the image, as the remaining part can be regarded as noise.
- The second stage is to extract relevant features from the skin segmented images which can prove significant for the next stage is learning and classification.
- The third stage is the extracted features as input into various supervised learning models for training and then finally use the trained models for classification.
- Then the CNN algorithm is used to extract and recognize the features of the hand gestures.
- We introduce a new technique called as Canny Edge Detection as it takes the images sharply by eliminating the background.
- In our proposed system we use some tools like Keras, Opencv, Tensor-flow, Webcam.

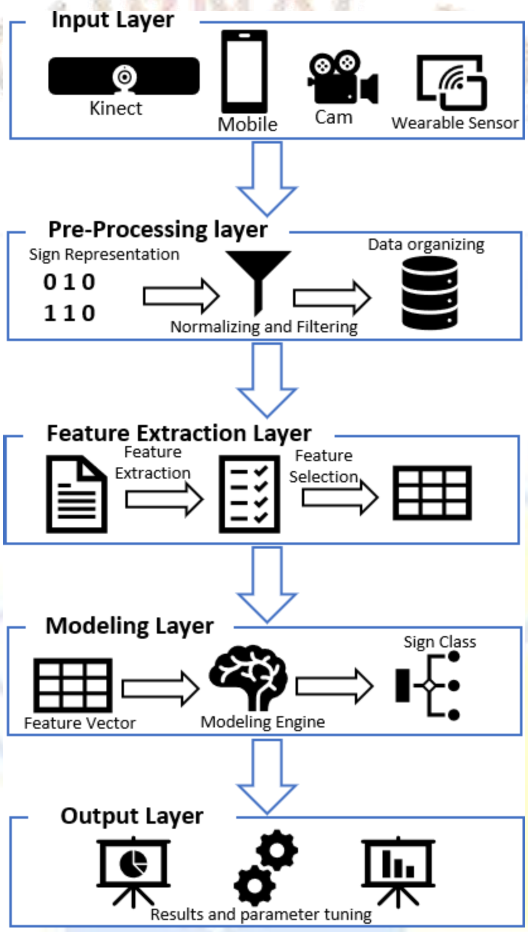

FIG 1. PROPOSED SYSTEM

IV. IMPLEMENTATION

The numerous techniques applied to the images in our dataset are described below.

- Grey Scaling
- Skin Masking
- Threshold
- Canny Edge Detection

Grey Scaling:

The first method used to process the images was by just converting the RGB image to Grayscale and resizing the image to pass through the CNN model used to train the images. Grayscale is a range of consistent monochrome shades from black to white. Digital images can be saved as or black and white images, even color images composed of grayscale information. Each pixel or intensity, which can be measured on a scale from black to white contains a luminance value, regardless of its color.

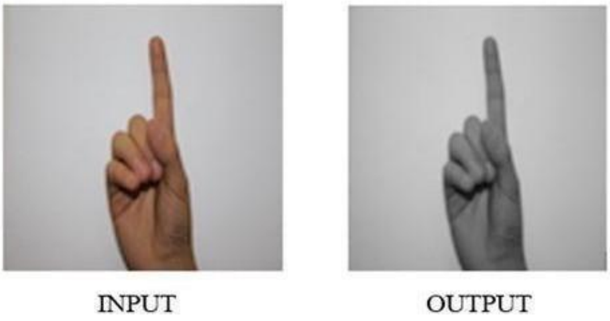

FIG 2. GRAYSCALE IMAGE

Skin Masking:

The first step of skin masking is to convert to grayscale as done. The next step is converting RGB to HSV. The images are converted to grayscale and then to HSV and are then given to the skin masking function. The color space of an image can be defined using HSV (Hue, Saturation, and Value). In HSV, the hue denotes the color. The hue is an angle from 0 degrees to 360 degrees. Saturation determines the range of grey in the color space.

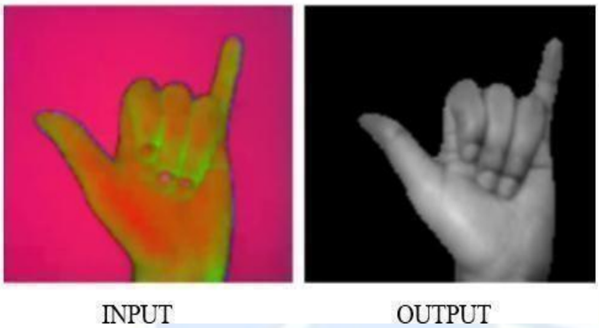

FIG 3. SKIN MASKING

Threshold:

Image thresholding is a simple way of dividing an image into the forefront and background. This image analysis technique is a type of image segmentation that divides objects by converting grayscale images into binary images. Image thresholding is effective to process an image with greater levels of contrast. The image is obtained after applying the above-mentioned threshold.

FIG 4. THRESHOL

Canny Edge Detection:

Canny edge detection is a technique to extract useful architectural information from different vision objects and adequately reduce the amount of data to be processed. The criteria for edge detection incorporate:

- The detection of edges should accurately capture as many edges shown in the image with a low error rate.
- An edge in the image should only be marked once, to avoid image noise and avoid the creation of false edges. Canny edge detection was applied to images after following the process of skin masking on the images to get defined edges of hand gestures and eliminating the background.

Convolution Neural Network:

- CNNs are a fundamental example of deep learning, where a more sophisticated model pushes the evolution of artificial intelligence by offering systems that simulate different types of biological human brain activity. The CNN model is fed with the processed images to classify the images.
- The method used to train the images. The CNN model with the processed images to classify the images. Once the model is fitted, predict random images from the computer to predict the text and display it on the screen.

- CNN architecture is implemented on the processed images. The results obtained by the various models are illustrated below.
- Various interpretations were used to measure the performance of the models. Before that here are the basic terms needed to understand-
- True Positives (TP) - The correctly predicted positive values which means that the value of actual class and the predicted class is true.
- True Negatives (TN) - The correctly predicted negative i.e., wrong values which means that the value of the actual class and predicted class is false.
- False Positives (FP) – When the actual class is false, and the predicted class is true.
- False Negatives (FN) – When the actual class is true, but the predicted class is false.
- Accuracy - Accuracy is the most accurate performance measure and it is a ratio of correctly predicted observation to the total observation.

V. CONCLUSIONS

The project is to remove the gap between users and the deaf, dumb people by introducing an economical computer application in the communication path so that the sign language can be automatically captured, recognized, and translated to text for the benefit of deaf people. The image obtained must be analyzed, processed and converted to either sign or textual display on the screen for the benefit of the hearing impaired. The system is an approach to ease the difficulty in communicating with those having speech disabilities. The amount of training and validation loss observed with the proposed CNN architecture was better. We have tried different image processing techniques to find the best one we need for our use. Firstly, we clone the required repositories for the process. Collecting the images using the label-Img API (Application Interface) to take the pictures using the real time webcam with the help of OpenCV. Label the images using label-Img app built with the help of Python libraries. Then train those images using the CNN (Convolution Neural Network). Update the check point of the trained images and detect the language in real time using webcam. To get more efficient results we used a technique named Canny Edge Detection (CED). During the live capture testing, the canny edge algorithm with 98% demonstrated results than the other techniques.

VI. REFERENCES

[1] T. Kim, J. Keane, W. Wang, H. Tang, J. Riggle, G. Shakhnarovich, D. Brentari, and K. Livescu, ‘‘Lexicon-free fingerspelling recognition from video: Data, models, and signer adaptation,’’ *Comput. Speech Lang.*, vol. 46, pp. 209–232, Nov. 2017.
[2] M. A. Ahmed, B. B. Zaidan, A. A. Zaidan, M. M. Salih, and M. M. B. Lakulu, ‘‘A review on systems-based sensory gloves for sign language recognition state of the art between 2007 and 2017,’’ *Sensors*, vol. 18, no. 7, p. 2208, 2018.
[3] R. Cui, H. Liu, and C. Zhang, ‘‘A deep neural framework for continuous sign language recognition by iterative training,’’ *IEEE Trans. Multimedia*, vol. 21, no. 7, pp. 1880–1891, Jul. 2019.
[4] P. S. Santhalingam, P. Pathak, J. Košecká, and H. Rangwala, ‘‘Sign language recognition analysis using multimodal data,’’ in *Proc. IEEE Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2019, pp. 203–210.
[5] A. C. Duarte, ‘‘Cross-modal neural sign language translation,’’ in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 1650–1654.
[6] M. J. Cheok, Z. Omar, and M. H. Jaward, ‘‘A review of hand gesture and sign language recognition techniques,’’ *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 1, pp. 131–153, Jan. 2019.
[7] Q. Xiao, Y. Zhao, and W. Huan, ‘‘Multi-sensor data fusion for sign language recognition based on dynamic Bayesian network and con-volutional neural network,’’ *Multimedia Tools Appl.*, vol. 78, no. 11, pp. 15335–15352, Jun. 2019.
[8] E. K. Kumar, P. V. V. Kishore, M. T. K. Kumar, and D. A. Kumar, ‘‘3D sign language recognition with joint distance and angular coded color topographical descriptor on a 2–stream CNN,’’ *Neurocomputing*, vol. 372, pp. 40–54, Jan. 2020.
[9] J. Wu and R. Jafari, ‘‘Wearable computers for sign language recognition,’’ in *Handbook of Large-Scale Distributed Computing in Smart Healthcare*. Cham, Switzerland: Springer, 2017, pp. 379–401.
[10] Q. Xiao, M. Qin, and Y. Yin, ‘‘Skeleton- based Chinese sign language recognition and generation for bidirectional communication between deaf and hearing people,’’ *Neural Netw.*, vol. 125, pp. 41–55, May 2020.

[11] W. Aly, S. Aly, and S. Almotairi, ‘‘User-independent American sign lan-guage alphabet recognition based on depth image and PCANet features,’’ *IEEE Access*, vol. 7, pp. 123138–123150, 2019.
[12] N. B. Ibrahim, H. H. Zayed, and M. M. Selim, ‘‘Advances, challenges and opportunities in continuous sign language recognition,’’ *J. Eng. Appl. Sci.*, vol. 15, no. 5, pp. 1205–1227, Dec. 2019.
[13] L. Zheng, B. Liang, and A. Jiang, ‘‘Recent advances of deep learning for sign language recognition,’’ in *Proc. Int. Conf. Digit. Image Comput., Techn. Appl. (DICTA)*, Nov. 2017, pp. 1–7.