2020 6th International Conference on Advanced Computing & Communication Systems (ICACCS)

Sign Language Translation

Harini R
Department of Computer Science and Engineering Saranathan college of engineering, Panjappur, Trichy, India.
hariniranga1998@gmail.com

Janani R
Department of Computer Science and Engineering
Saranathan college of engineering, Panjappur, Trichy, India

Keerthana S
Department of Computer Science and Engineering
Saranathan college of engineering, Panjappur, Trichy, India

Madhubala S
Department of Computer Science and Engineering
Saranathan college of engineering, Panjappur, Trichy, India

Venkatasubramanian S
Department of Computer Science and Engineering
Saranathan college of engineering, Panjappur, Trichy, India

ABSTRACT-- **Sign language is the way of communication for hearing impaired people. There is a challenge for common people to communicate with deaf people which makes this system helpful in assisting them. This project aims at implementing computer vision which can take the sign from the users and convert them into text in real time. The proposed system contains four modules such as: image capturing, pre-processing classification and prediction. By using image processing the segmentation can be done. Sign gestures are captured and processed using OpenCV python library. The captured gesture is resized, converted to grey scale image and the noise is filtered to achieve prediction with high accuracy. The classification and predication are done using convolution neural network.**

KEYWORDS -- Machine learning; Gesture recognition; Sign language; Human Computer Interaction; Computer Vision

I. INTRODUCTION

The aim of this system is to elevate people with hearing disability and help them socialize with common people. It is a form of non-verbal communication. Sign language is the structured form as each gesture represents a unique element or a character. With the advent of advancement in science and engineering many researchers are working on different methodologies that could take the event of human computer interaction to a much higher extent. The computer is trained in such a way that it could translate the sign to text for static as well as dynamic frames. The system is designed and implemented to recognize sign language gestures. The signs are captured using web cam and are pre-processed. In pre-processing stage, we use background subtraction to eliminate the background which makes this system to adapt to any dynamic background. The main difficulty while implementing in software based is that the image must be properly captured and filtered.

II. LITERATURE SURVEY

The related work on this project shows that there have been several methods of implementing the system under different domains namely vision-based approach, glove-based approach, fuzzy logics, soft computing like neural network or using MATLAB etc. Vision-based approach requires camera to capture image in 2D or 3D format. [1] in this proposed system canny edge detection algorithm was used as the accuracy of this algorithm is better and the time consumed is also less. This algorithm is efficient in removing noise as well as detecting a clear image from the input for further processing. This algorithm results comparatively a low rate of error, edge points that are localized, and single edge point response. This system has been implemented using java NetBeans.

[2] This system is designed and built in such a way that the gestures are detected from the obtained input images and are converted to its standard format i.e. the converted hand gestures are represented in English language. The input that is in the form of video is segmented into single frames and each frame is then given to its pre-processing function. In order to reduce the unwanted regions and to enhance its performance each frame travels through multiple filters. Fourier description technique is adapted to extract hand gestures from the pre-processed frames and are finally stored in its desired repository

[3] This system is based on the supervised machine learning called Support Vector Machine (SVM). The authors of this paper have gathered dataset from students who has a prior knowledge or are gone through training to perform sign gestures. The dataset was captured using a still camera of 1.3M pixels. The accuracy for this system after training was found to be cent percent since only a few gestures were taken under account.

978-1-7281-5197-7/20/$31.00 ©2020 IEEE

[4] This system uses a new methodology to train on the Haar features of the image based on a classifier named AdaBoost. This AdaBoost helps in detecting left or right hand at a time and combines Haar-like features and frame difference to learn the skin-color. The classifier also gives an extended tracking method by which the hand patch of the last frame is captured and processed so as to produce a new patch in current frame. The rate at which the system predicts the gestures is found to be 99.9%.

[5] This system is based on perceptron- a mechanism of neural networks. The design is developed with the help of SciPy- an inbuilt python package. The performance metrics of the system uses few parameters such as precision, recall, F1 score. In order to improve performance, the system uses Pruning technique which helps to trim the network size. During the training phase the hidden layers were gradually incremented starting from 10 till 120.

[6] In this system, the sign recognition is implemented using ARM CORTEX A8 processor. OpenCV python library is used for image capturing and image processing in real time. To predict on both positive as well as non-positive images Haar training features were used.

[7] This system is trained with dataset containing 300 different images of ISL numbers. The accuracy of the system was 99.56% in 22 epochs. The system was experimented with different learning rates and activation functions. Keras API with Tensor Flow is used as backend. The system has predicted the static symbols accurately when tested with 100 images for each sign.

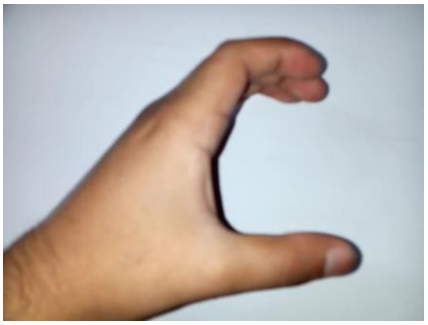
[8] In this system, four convolutional layers with different window sizes along with ReLu activation function were used. The system was tested with three pooling strategies in which stochastic pooling is found to be effective. Two layers of stochastic pooling were used for feature extraction.

[9] To extract spatial features, this system used Inception model of CNN for sign gesture recognition. To extract temporal features from the video streams LSTM (Long Short-Term Memory) and RNN (Recurrent Neural Network) model were used. The CNN and RNN were evaluated independently using the same training and test samples. This ensures that neither CNN nor RNN uses the test data during the training phase to achieve better prediction. To minimize loss, Cross-entropy cost function ADAM optimizer was used to train both the models.

[10] The proposed system was implemented using a 2-layer convolutional neural network (CNN) to predict sign language gestures made by the users. Two different models were used to classify and compare the accuracy of prediction. The two optimizers used for optimizing the output were SGD optimizer and Adam optimizer, in which Categorical Cross entropy cost function was used. The model was found to predict the gestures well even with noisy images and in different lighting conditions of datasets. Using SGD and Adam as two different optimizers, this system has classified 6 different sign languages with an accuracy of 99.12% and 99.51% respectively. More accuracy is obtained when the Adam optimizer is used.

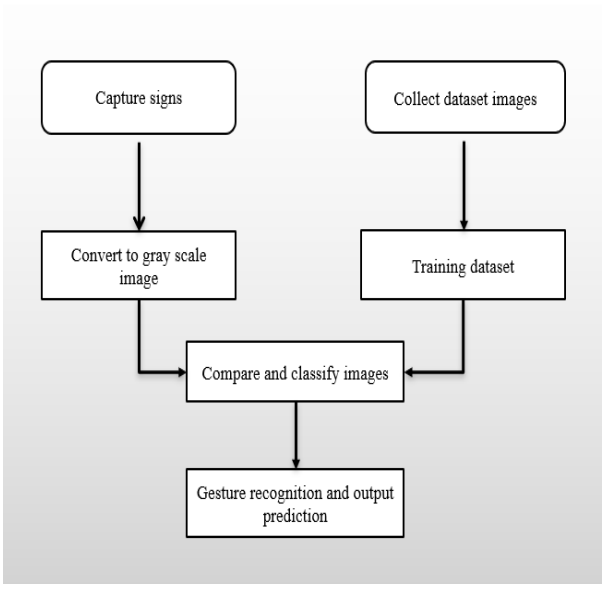
III. DATASET

We have created dataset containing thousands of images of each category and converted the images into a CSV file so as to get speed and accuracy in training the system. For rare gestures we tried to capture the signs using our mobile camera. The captured images were converted to desirable size and pre-processed for classification. The captured image is converted to its pixel values and stored in csv file format.



IV. PROPOSED SYSTEM

The foremost aim of our system is to provide communication between common people and those with hearing aid without need of any specific color background, hand gloves or any sensors. Other systems used image dataset as such in '.jpg' format. But in our system the pixel values of each images are stored in csv file which reduces the memory requirement of the system. Also, the accuracy of prediction is high when csv dataset is used. The four modules of this system are Image capturing, Preprocessing, Classification and Prediction.



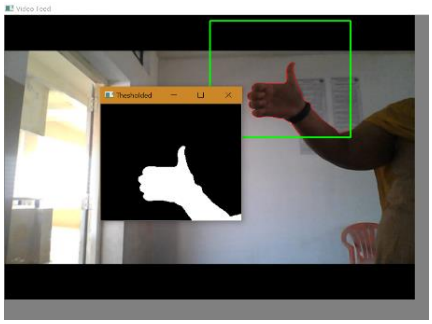
A. IMAGE CAPTURING

Python OpenCV library can be used to capture sign gestures from computer's internal camera. The dataset for various signs is collected. To predict gestures with high accuracy, around 3000 images are collected for each sign.

The collected dataset images are converted to csv file containing the pixel values for higher accuracy.

B. PRE-PROCESSING

The primary focus of the system is to support detecting gestures in dynamic background condition. To achieve this, the frames are preprocessed and converted to gray scale image and then background subtraction algorithm is used. The camera first captures around 10 frames to identify the background and compares current frame with previous frame. If a hand is detected then the background is subtracted and only the hand gesture is converted to gray scale. Later the converted hand portion is resized to 28*28 size to extract features.



C. CLASSIFICATION

After collecting and processing the image dataset, they have to be classified. Convolutional neural network is used to analyze and classify visual imagery. It is widely applied in image recognition and classification, natural language processing, etc. CNN is regularized versions of multilayer perceptron. It consists of convolutional layer, pooling layer, flattening and fully connected layer along with activation functions.

- *CONVOLUTION LAYER*

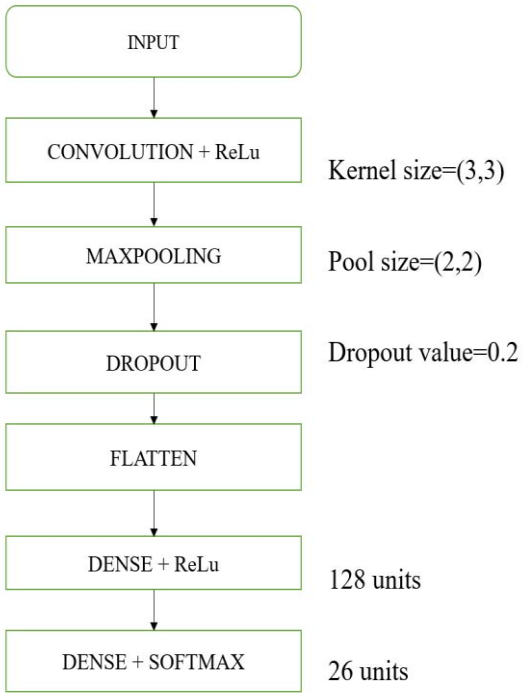
It is performed on an image to identify certain features in an image. This layer accepts the input signal and filters it for feature extraction. Convolution layer uses image matrix and kernel matrix to extract features from image. The output is dot product of image matrix and kernel matrix.

- *POOLING LAYER*

A convoluted image can be too large or too small and therefore needs to be reduced without losing its features. Two types of pooling are max and min pooling. In Max pooling layer the maximum value is selected from the particular region and in min pooling layer the minimum value from the particular region is selected.

- *FLATTENING LAYER*

This layer transforms multi-dimensional matrix to 1-d array so that it can be fed into the classifier.



- *ACTIVATION FUNCTION*

The activation function decides whether a neuron has to be activated or not based on its output. There are different activation functions like sigmoid, tanh, ReLu, etc. In our system the activation function used is ReLu. It is the most widely used activation function. If the input of any neuron is negative then it is considered as 0 else 1. Because of this, the accuracy of ReLu is high and computation is simple. The ReLu activation function can be represented as **f(x) = max (0, x).** i.e

$$f(x)= \begin{cases} 0, \text{if } x<0 \\ x, \text{if } x>0 \end{cases} \quad (1)$$

V. CONCLUSION

The proposed system translates the sign gestures to text using CNN. The accuracy of the model is 99.91%. Although the facial expressions express a lot during communication, the system does not focus on facial expressions. The accuracy of the model was less with poor lighting. As future enhancements, more dynamic video signs can be trained involving facial features and expressions.

REFERENCES

[1] Amitkumar Shinde, Ramesh Kagalkar," Sign Language to Text and Vice Versa Recognition using Computer Vision in Marathi," National Conference on Advances in Computing (NCAC 2015).

[2] Purva C. Badhe, Vaishali Kulkarni," Indian Sign Language Translator Using Gesture Recognition Algorithm," IEEE International Conference on Computer Graphics, Vision and Information Security (CGVIS 2015).

[3] Rajesh Mapari, Govind Kharat ," Hand gesture recognition using Neural Network,December," International Journal of Computer Science and Network. 2012

[4] S. Wu and H. Nagahashi, "Real-time 2D hands detection and tracking for sign language recognition,"8th International Conference on System of Systems Engineering. 2013

[5] Geethu G Nath, Anu V S," Embedded Sign Language Interpreter System for Deaf and Dumb People," International Conference on Innovations in information Embedded and Communication Systems (ICIIECS). 2017

[6] Sajanraj T D, Beena M V," Indian Sign Language Numeral Recognition Using Region of Interest Convolutional Neural Network," 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT 2018).

[7] G.Anantha Rao, K.Syamala, P.V.V.Kishore, A.S.C.S.Sastry,"Deep Convolutional Neural Networks for Sign Language Recognition"

[8] Kshitij Bantupalli, Ying Xie, "American Sign Language Recognition using Deep Learning and Computer Vision," IEEE International Conference on Big Data 2018(Big Data).

[9] Surejya Suresh, Mithun Haridas.T.P, Supriya M.H." Sign Language Recognition System Using Deep Neural Network,"5th International Conference on Advanced Computing & Communication Systems (ICACCS) 2019.