INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 7, Issue 4 - V7I4-1141)
Available online at: https://www.ijariit.com

Sign Language recognition system using Convolutional Neural Networks

Puramshetty Srinidhi Rao
puramsrinidhi10@gmail.com
Andhra University College of Engineering for Women, Visakhapatnam, Andhra Pradesh

Potladurthy Sindhu Sri Vatsalya
sindhussv124@gmail.com
Andhra University College of Engineering for Women, Visakhapatnam, Andhra Pradesh

Ruttala Bhanu Sree Sai
bhanuramam1108@gmail.com
Andhra University College of Engineering for Women, Visakhapatnam, Andhra Pradesh

Polavarapu Janshi
p.jhansi1999@gmail.com
Andhra University College of Engineering for Women, Visakhapatnam, Andhra Pradesh

Koda Indu
indu.gf@andhrauniversity.edu.in
Andhra University College of Engineering for Women, Visakhapatnam, Andhra Pradesh

ABSTRACT

Communication is key to share and express ideas and thoughts. And not every person has the natural ability to communicate. Mute people communicate with each other and with rest of the people through sign language. Sign language is the way of communication using gestures with hands where each gesture corresponds to particular letter in an alphabet or a word. But to use this every person must have the knowledge of it. But in real world, it is not the case. With the advancement in the field of deep learning it is possible to solve this problem with low cost. So, this project aims at making it easier for mute people to communicate to the rest of the world. In this project, image of the hand gesture made by the person is captured and is translated to its corresponding meaning in the American sign language system. Equipment such as hand gloves sensor is one of the methods used for sign language detection. But it is expensive and not all of the people in need of it would be able to afford it. A deep learning algorithm called convolutional neural network is used for identifying the gestures and giving the actual letter or word.

Keywords: *Sign language Detection, Image Processing, Convolutional Neural Network.*

1. INTRODUCTION

Sign language is the mode of communication which uses visual ways like expressions, hand gestures, and body movements to convey meaning. Sign language is extremely helpful for people who face difficulty with hearing or speaking. Sign language recognition refers to the conversion of these gestures into words or alphabets of existing formally spoken languages. Thus, conversion of sign language into words by an algorithm or a model can help bridge the gap between people with hearing or speaking impairment and the rest of the world. If there is a common interface that converts the sign language to text the gestures can be easily understood by the other people. The sign language recognition architecture can be categorized into two main classifications based on its input: data gloves-based and vision-based. Using smart gloves to acquire measurements such as the positions of hands, joints, and velocity using microcontrollers and specific sensors. There are other approaches to capturing signs by using motion sensors, such as electromyography sensors, leap motion controllers or their combinations. The advantage of this approach is having higher accuracy, and the weakness is that it has limited movement. In recent years, the involvement of vision-based techniques has become more popular, of which input is from camera (web camera, stereo camera, or 3D camera). A combination of both architectures is also possible, which is called the hybrid architecture. While these are more affordable and less constraining than data gloves, the weakness of this approach is lower accuracy and high computing power consumption. So, research has been made for a vision-based interface system where deaf and mute people can enjoy communication without really knowing each other's language. The basic task in this project is to recognize the gesture of the hand and predict which letter it corresponds to using the trained model. The image captured is subjected to image processing methods like changing the image from one colour space to the other, removing high frequency noise using the gaussian filter, reducing noise using the median filter, performing shape analysis by applying contour filters. And then by using the convolutional neural networks the classification is performed. The English alphabet and numbers are

trained. The aim is to develop a user-friendly human computer interfaces where the computer understands the human sign language.

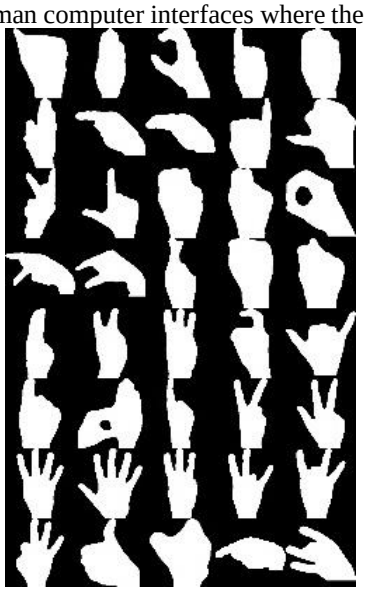

Chart 1: Gestures used in the project

2. EXISTING SYSTEM

The previous methods in sign language recognition used electrical equipment like the smart gloves and sensors which is a costly affair. And the previous methods had high latency because of the insufficient training of the model and pictures.

3. PROPOSED SYSTEM

In the proposed system, we train the model using 1200 images of each gesture of the American sign language system. We use methods from the python opencv library for the purpose of image processing which can make training and recognition part easier. Convolutional neural network algorithm is applied for the classification of the image. This decreases the latency of the project and improves the accuracy. It would be a real time system where live image is processed by taking the input from the webcam. It provides more interactivity than the existing system.

4. WORKING PROCEDURE

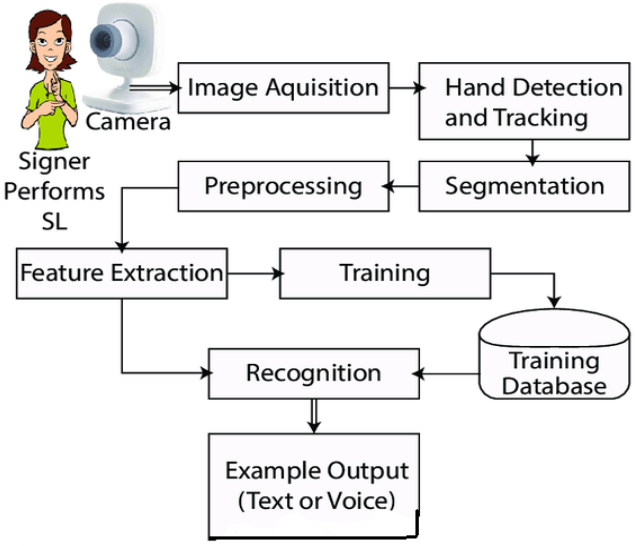

Chart 2: Architecture of the System

4.1 Algorithm

Step1: Data Collection Image capture with web camera
Step 2: Image Processing
- Backgrounds detected and eliminated with HSV
- Morphological operations are performed and masks are applied.
- Segmentation of hand gesture
- Image is resized

Step 3: Extracting the features from the processed images
Step 4: Creating a CNN model to train
Step 5: Training the model by giving the images as input
Step 6: Testing the model to verify the model and calculate the accuracy

4.2 Data Collection

The model is first trained by using gestures of the English alphabet and the numbers. It is trained by using 1200 pictures of each gesture and then assigning a value to each gesture. It is a supervised learning method. The users and the background are not saved in training data set. The image is saved after undergoing preprocessing. The background in the test set can occur in the testing and validation set.

4.3 Processing of the input image

While training the model and while classifying the gestures, the image captured through the camera is applied with some preprocessing techniques for making the process easier for the system. The picture captured is first changed from RGB color space to the HSV color space. Now the high frequency components are removed from the picture by using gaussian filter. The median blur filter is applied where the pixels of the frame are replaced by the median the surrounding pixels. This is used to reduce the noise in the image. Thresholding is applied to change the image from grayscale to binary to make the analyzing part easier. The contours of the image are detected using the methods for the purpose of shape analysis. Finding the contours is a type of feature extraction method. Feature extraction is a very useful step. After performing all the image processing the image is converted into black and white image with the required features.

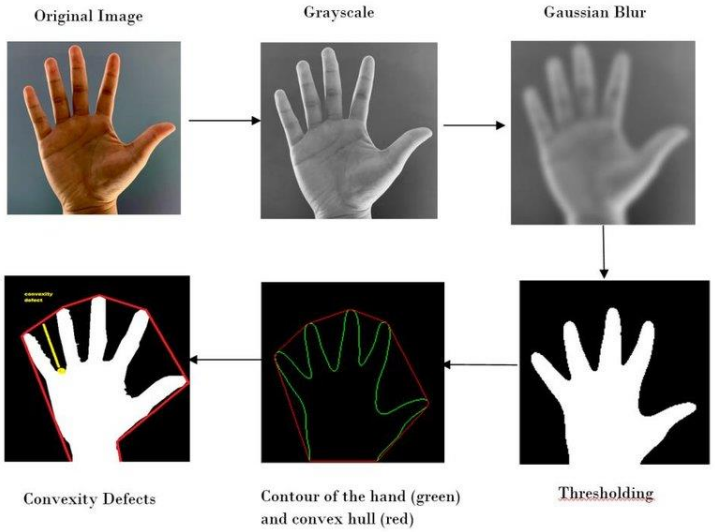

Chart 3: Preprocessing of the captured image

4.4 Convolutional Neural Network

A convolutional neural network is a deep learning algorithm which can take in image, assign importance that is learnable weights and biases to various aspects in the image and be able to differentiate from one another. A convolutional neural network is an improvement it of artificial neural network. Say we have a simple digit or alphabet. It is represented as a matrix where each pixel value is RGB number. Now that grid is flattened to make a single array. It is then it adjusts the weights of node to get the accurate output. But this doesn't work on a bigger picture where the total number of nodes would be around a few millions. In CNN, we extract features of the image by using filters. The matrix of the filter is multiplied to the matrix of the pixel grid of the image to get a feature map. The image is multiplied with multiple feature maps then we take the feature map and perform ReLu operation that is non-linearity operation. It replaces negative values with positive values. Pulling operation is then used to reduce the size of the the Matrix. In this project we used Max pooling operation. The two operations of convolution and ReLu operation and pooling is performed multiple Times. This is called feature extraction. This is the first part of CNN. The final image so obtained it is much lesser then the one in the artificial neural network. Now this grid of matrix is flattened to a single array. This is the input layer of the neural network. There will be many hidden layers in between. The model then assign weights to neurons to get the corresponding value of the gesture. This part it is called classification. In this way the real time image is converted to its corresponding letter or digit.

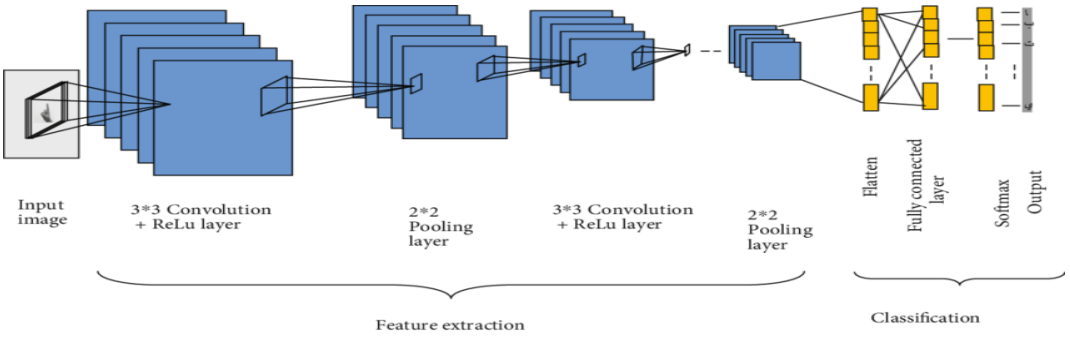

Chart 3: Convolutional neural network

5. CONCLUSION

The main objective of the project was to develop a system that can translate static sign language into its corresponding word equivalent that includes letters, numbers, and basic static signs to familiarize the users with the fundamentals of sign language. This

work shows that convolutional neural networks can be used to accurately recognize different signs of a sign language, with users and surroundings not included in the training set. This generalization capacity of CNNs can contribute to the broader research field on automatic sign language recognition. Despite it having average accuracy, our system is still well-matched with the existing systems, given that it can perform recognition at the given accuracy with larger vocabularies and without an aid such as gloves or hand markings.

6. REFERENCES

[1] P. Mekala, Y. Gao, J. Fan, and A. Davari, "Real-time sign language recognition based on neural network architecture"
[2] C. Vogler and D. Metaxas, "Handshapes and movements: Multiple-channel american sign language handshapes and movements: Multiple-channel ASL recognition"
[3] Q. Wang, X. Chen, L. G. Zhang, C. Wang, and W. Gao, "Viewpoint invariant sign language recognition"
[4] I. N. Sandjaja and N. Marcos, "Sign language number recognition"
[5] T. Chouhan, A. Panse, A. K. Voona, and S. M. Sameer, "Smart glove with gesture recognition ability for the hearing and speech impaired,"
[6] C. Ong, I. Lim, J. Lu, C. Ng, and T. Ong, "Sign-language recognition through Gesture & Movement Analysis (SIGMA),"
[7] S. Shahriar, A. Siddiquee, T. Islam, A. Ghosh, R. Chakraborty, A. I. Khan, C. Shahnaz, and S. A. Fattah, "Real-time american sign language recognition using skin segmentation and image category classification with convolutional neural network and deep learning,"
[8] R. Daroya, D. Peralta, and P. Naval, "Alphabet sign language image classification using deep learning,"
[9] T. Kim, G. Shakhnarovich, and K. Livescu, "Finger-spelling recognition with semi-markov conditional random fields,"
[10] N. Pugeault and R. Bowden, "Spelling it out: Real-time asl finger-spelling recognition,"

