International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 05 | May - 2022 Impact Factor: 7.185 ISSN: 2582-3930

# Sign Language Recognition and Interpretation System

Sakshi Gaye*, Vaishnavi Patil**, Keshav Rai***, Anurag Parate ****

(Department of Information Technology, Government College of Engineering, Amravati – 444604, Maharashtra, India)

## I. ABSTRACT

Sign language is a visual language that individuals with speech and hearing difficulties use to communicate in their daily conversations. Through its original grammar, it is totally an optical communication language, as opposed to vocal languages. This study report proposed an effective strategy to accomplishing the translation of 24 static sign language alphabets and numerals of American Sign Language into humanoid. Researchers are now focusing their efforts on building commercially viable Sign Language Recognition systems. Researchers use a range of techniques to perform their research. It all starts with the methods of data acquisition. The data collection method varies depending on the cost of a suitable device, however for the Sign Language Recognition System to be commercialized, a low-cost strategy is required. The methods used by researchers to develop Sign Language Recognition varied as well. Each approach has its own advantages over others, and researchers are continually experimenting with various methods to produce their own Sign Language Recognition. Each approach has its own set of limitations when compared to others. The purpose of this study is to examine several ways to sign language recognition and choose the optimal method.

***Keywords* —** Feature Extraction and Representation, Artificial Neural Networks, Convolution Neural Network, TensorFlow, Keras, OpenCV.

## II. INTRODUCTION

Instead of acoustic sound patterns, Sign Language is a gesture-based language that involves hand movements, hand orientation, and face expression. This language has different patterns depending on the people and is not universal.

Deafness is a disability that impairs hearing and causes them to be unable to hear, whereas mutism is a disability that impairs speech and causes them to be unable to talk. Both are only deaf or deafeningly deafening Communication is the single thing that separates them fromordinary people. If normal people and deaf-mute people can communicate, the deaf-mute people can easily live regular lives. And their only means of communication is through sign language.

Sign languages are not standard or universal, and their grammars differ from one another. For example, hand movements, body movements, and facial expressions are all involved in Portuguese Sign Language (PSL). The goal of Sign Language Recognition (SLR) systems is to provide an efficient and accurate means to convert sign language into text or speech for use as aids for the hearing impaired, or to enable very young children to engage with computers (recognizing sign language), among other things.

The purpose of this study is to discuss the Sign Language Recognition System that researchers employ. In this paper, we will look at Sign Language Recognition from application point of view.

## III. RELATED WORK

Sign languages are defined as an organized collection of hand gestures with specific meanings used by hearing impaired people for day-to-day communication. Over 300 different sign languages are available all over the world. Despite the fact that there are numerous sign languages, the percentage of the community who understands any of them is minimal, making it difficult for persons with disabilities to communicate freely with everyone.

Kalidolda's [1] paper presented the research, which attempted to create a sign language interpreting robotic system. The project had two main goals: to facilitate real-time fingerspelling recognition and to conduct performance testing "in the wild," as well as to gather input on the NAO robotic translating system from deaf-mute people. The robotic system is made up of several software and hardware components, including the Microsoft Kinect SDK for person identification and tracking and the Leap Motion SDK for hand detection and tracking. Serving as a social interface is a stationary humanoid NAO robot (NaoQi SDK). NAO may potentially be used as a robotic instructor for kids. As a social interface, a stationary humanoid Pepper robot (Pepper SDK for Android). For teaching apps, a computer monitor and an Android tablet (Android SDK) are used.

Jiyong [3] was used as an input to Hidden Markov Models (HMMs) for a real-time system meant to recognize continuous Chinese Sign Language (CSL). Two Cyber- Gloves and a 3-D tracker provided raw data. The Welch- Baum algorithm was used to estimate after segmenting the training text into basic units using the dynamic programming(DP) technique. The system achieved a 94.7 percent recognition rate when tested with 220 words and 80 sentences.

Volger [2] employed Parallel Hidden Markov models (PaHMMs) to recognize American Sign Language. They claimed that phonemes, rather than complete signals, might be recognition, can be broken down into fundamental phonemes. A single channel of the HMM model was evaluated for a limited vocabulary number (22 signs), with an accuracy rate of 87.88 percent. A greater vocabulary, hand arrangement, orientation, and facial emotions are not recognised by the system.

An SLR system was created using a variety of processing approaches. The Hidden Markov Model (HMM) is commonly utilized in SLR. Multi Stream HMM (MSHMM), which is based on the two-standard single-stream HMMs, Light-HMM and Tied-Mixture Density MM, has been employed. Other processing models used include neural networks, ANN, Nave Bayes Classifier (NBC), and Multilayer Perceptron (MLP), unsupervised neural networks, Self-Organizing Map (SOM), Self-Organizing Feature Map (SOFM), Simple Recurrent Network (SRN. Researchers have also developed their own approaches, such as the wavelet-based method and Eigen Value Euclidean Distance.

The adoption of various processing methods or application systems has resulted in varying degrees of accuracy. The Light-HMM achieved 83.6 percent accuracy, the MSHMM achieved 86.7 percent accuracy, the SVM achieved 97.5 percent accuracy, the Eigen Value achieved 97 percent accuracy, and the Wavelet Family achieved 100 percent accuracy. Though several models have produced high accuracy results, the accuracy is dependent on a variety of criteria, including the size of the dataset, the clarity of the images of the dataset based on data gathering methods, equipment employed, and so on.

## IV. METHODOLOGY

Sign Language, like any other language used by the deaf people, is a form of communication. This dataset contains acomplete collection of sign language motions that can be utilizedby other normal individuals to improve their knowledge of sign language gestures. This data is entirely made up of classes that contain all of the English alphabet's characters. The Model is written in Python, which is a common programming language forML and DL tasks.VPytorch is utilized in this project because of its customizable library and features. More information on the models and others is provided below.

The ML Life cycle is divided into four major stages.

- Acquisition of Data
- Preprocessing and extracting features from data
- Model Preparation
- Deployment of the Model

The system is a vision-based approach. All of the signs are

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 05 | May - 2022 Impact Factor: 7.185 ISSN: 2582-3930

portrayed with bare hands, there is no need for any artificial gadgets for interaction.

### *A.* DATA SET GENERATION

. The steps we took to construct our data set are shown below. Following the collection of the dataset, the classification issue is broken down into three parts.

• The first step is to segment the skin part from the image, as the remaining part can be regarded as noise with respect to the character classification problem.

Fig. 1 Capturing Raw Image

• The second step is to extract relevant features from the skin segmented images which can prove significant for the next step i.e., learning and classification.

Fig. 2 Gray Scale Image

• The third step as mentioned above is to use the extracted features as input to various supervised learning models for training and then finally trained models can be used for classification.

Fig. 3 Image Post Gaussian Blur

### *B.* GESTURE CLASSIFICATION

Our strategy for this project was to forecast the user's final symbol using two levels of algorithms.

**Algorithm Layer 1:**

1. Apply gaussian blur filter and threshold to the frame taken using OpenCV to get the processed image after featureextraction.
2. This processed image is then passed to the CNN model for prediction and if a letter is detected for more than 50 frames then the letter is printed and taken into consideration for forming the word.
3. Space between the words is considered using the blank symbol.

**Algorithm Layer 2:**

1. We detect various sets of symbols which show similar results on getting detected.
2. We then classify between those sets using classifiers made for those sets only.

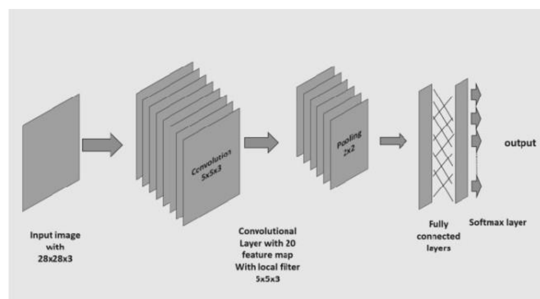


Fig.4 Convolutional Neural Network

### *C.* FINGER SPELLING SENTENCE FORMATION IMPLEMENTATION

1. When the count of a detected letter exceeds a certain threshold and no other letter is within a certain distance of it, the letter is printed and added to the current string (In our code we kept the value as 50 and difference threshold as 20).
2. Otherwise, we clear the current dictionary, which contains the number of detections of the current symbol, to reduce the possibility of predicting the wrong letter.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 05 | May - 2022 Impact Factor: 7.185 ISSN: 2582-3930

3. When the count of detected blanks (plain backgrounds)

exceeds a certain threshold and the current buffer is empty,no spaces are recognised.

4. In the other situation, it forecasts the end of the word byprinting space, and the current sentence is appended below.

### *D.* ARCHITECTURE

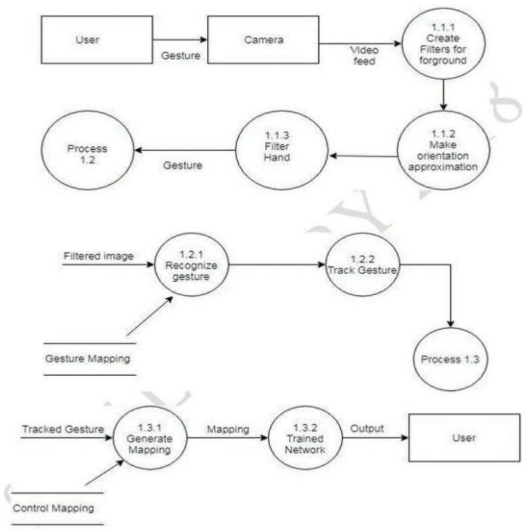


Image data has to be converted into an array list so they canbe further used in the machine learning model.

For the conversion, we grab the list of images in the dataset and initialize them using TensorFlow functions.

We have converted the images to a (224,224) pixel array. Then the data is preprocessed i.e., it is converted into a formthat could be provided to ML algorithms to do a better job inprediction.

## V. RESULTS

We attained an accuracy of 95.8 percent in our model using only layer 1 of our method, and an accuracy of 98.0 percent when layer 1 and layer 2 were combined, which is higher than the majority of current research articles on American sign language. The majority of the research publications concentrate on employing devices such as Kinect for hand detection. In, they use convolutional neural networks and Kinect to create a recognition system for Flemish sign language with a 2.5 percent mistake rate. A recognition model is developed with a hidden Markov model classifier and a lexicon of 30 words, and the error rate is 10.90%. Japanese sign language, they attain an average accuracy of 86 percent for 41 static movements. Map obtained an accuracy of 99.99 percent for observed signers and 83.58 percent and 85.49 percent for fresh signers using depth sensors. CNN was also employed in their recognition system. It should be emphasized that our model does not employ any background subtraction algorithms, but several of the models presented above do. As a result, the accuracy of our project's background subtraction may vary. On the other hand, while the majority of the above 21 projects make use of Kinect devices, our main goal was to designa project that could be used with widely available resources. A sensor like Kinect is not only not widely available, but also prohibitively expensive.

## VI. FUTURE SCOPE

1. This approach can be applied to other sign languages, such as Indian Sign Language, however it is currently limited to American Sign Language.

2. The model may then be trained with a dataset to automatically segregate the gesture from the collected frame by deleting the backdrop.

3. The dataset can also be supplemented with words and short sentences such as "What's on your mind?"

4. Model tuning and enhancement to recognize common words and expressions

5. Additionally, training the neural network model to recognize symbols that require two hands.

6. We intend to improve accuracy even in the presence of complicated backdrops by experimenting with various background subtraction algorithms. We are also considering ways to improve the preprocessing to predict gestures in low light conditions with a higher accuracy.

## VII. CONCLUSION

For asl alphabets, a functional real-time vision-based American sign language recognition system for D&M individuals has been developed. On our dataset, we attained a final accuracy of 98.0 percent. We can improve our prediction by developing two layers of algorithms that check and predict symbols that are more similar to one another. This method allows us to recognize practically any symbols as long as they are adequatelydisplayed, there is no noise in the background, and the illumination is acceptable.

## VIII. ACKNOWLEGEMENT

Dr. P. B. Ambhore, Dept. of Information Technology (I.T.), Government College of Engineering, Amravati, has been a source of support and guidance throughout this research.
We are also obliged to our Head of Department (Dept. of Information Technology), Prof. A. W. Bhade for giving us this great opportunity. We are thankful for their cooperation during the period of our assignment and research.

## XI. REFERENCES

[1] Kalidolda's, N., &Serdyukov, A. (2018). Towards Interpreting Robotic System for Finger spelling Recognition in Real-Time. HRI Companion:2018 ACM/IEEE International Conference on the Human-Robot Interaction Companion.

[2] Vogler, C.M., &Dimitris (2004). Handshapes and Movements: Multiple-Channel American Sign Language Recognition Gesture-Based Communication in Human-Computer Interaction. Ed. A.V. Camurri, Gualtiero. Vol. 2915: Springer Berlin Heidelberg. 247-258.

[3] Jiyong, M.W., Gao, Jiangqin, &Wu Chunli, Wang (2000). A Continuous Chinese sign language recognition system in Automatic Face and Gesture Recognition. Fourth IEEE International Conference on 2000.

[4] Gunasekaran, K., &Manikandan, R. (2013). Sign Language to Speech Translation System Using PIC Microcontroller. International Journal of Engineering and Technology (IJET), Vol 5 No 2.

[5] Liang, R., & Ouhyoung, M. (1998). Real-time continuous gesture recognition system for sign language. Proc Third. IEEE International Conf: on Automatic Face and Gesture Recognition, pp. 558-567.

[6] Kalsh EA, Garewal NS. Sign Language Recognition System. International Journal of Computational Engineering Research. 2013; 03(6): p. 15- 21.

[7] Tewari D, Srivastava S. Based on a visual recognition of static hand gestures in Indian Sign Language Kohonen Self-Organizing Map Algorithm. International Journal of Engineering and Advanced Technology (IJEAT). 2012; 2(2): p. 165-170.

[8] Raheja JL, Mishra A, Chaudary A. Indian Sign Language Recognition Using SVM 1. Pattern Recognition and Image Analysis. 2016 September; 26(2).

[9] Kishore PVV, Prasad MVD, Prasad CR, Rahul R. 4-Camera model for sign language recognition using elliptical fourier descriptors and ANN. In International Conference on
Signal Processing and Communication Engineering Systems - Proceedings of SPACES 2015, in Association with IEEE; 2015. p. 34-38.

[10] Goyal, Sakshi & Sharma, Ishita & Sharma, Shanu. For Deaf and Dumb People, A Sign Language Recognition System Engineering Research & Technology: An International Journal (IJERT). 2013 April; 2(4).

[11] Huang J, Zhou W, Li H, Li W. Sign language recognitionusing 3D convolutional neural networks. In Multimedia and Expo (ICME), 2015 IEEE International Conference on; 2015: IEEE. p. 1-6.