WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

An Analysis of Sign Language Translation and Recognition Systems

P. Pallavi[a], A. N. V. Sandeep[a], B. Amman Naidu[a], G. Tirupathi[a], D. Praharsha Reddy[a], Ch. Anil Kumar[b]

[a]Student, [b]Assistant Professor
Department of Information Technology, GMR Institute of Technology,
Rajam – 532127, Andhra Pradesh, India.
DOI: https://doi.org/10.55248/gengpi.2022.3.4.9

A B S T R A C T

Communication is very essential for each and every individual person, because it allows us to express our feelings, sentiments, and ideas etc. Humans can communicate with one another through a variety of means, including natural language, body language, and even writing. People who do not have any disabilities, such as hearing or speaking impairments, can use all of the foregoing modes of communication to communicate with other normal people. Natural language, which includes words, sentences, and so on, is the most often utilized type of communication by normal people. On the other hand, People with disabilities use Sign Language to communicate with others. Deaf and dumb people comprehend sign language easily and quickly, but it is difficult for them to communicate with normal people who do not understand sign language. Deaf and dumb persons are unable to effectively convey their feelings and thoughts to other normal people due to hearing and speaking issues, which is a serious problem all over the world. To address this issue, numerous systems were developed that interpret sign language (Hand movements, facial expression etc.), recognize signs (Hand Gestures), and display the relevant text on the system, allowing normal people to understand the language of the deaf and dumb. The fundamental purpose of this survey paper is to read all of the methods that were used in various systems and to identify the most efficient method that can be adopted in future research work.

Keywords:Sign Language, Impaired Persons, Hand Gestures, Text etc.

1. Introduction

The WHO (World Health Organization) conducted survey/research in the year 2021 and discovered that approximately 5-10% of the world's population suffers from hearing and speaking impairments. India is the world's second-largest country, with around 63 million people suffering from deafness or dumbness. These people interact with normal people in a variety of methods, including writing, text messaging, sign language, and so on. However, Sign language is widely considered as one of the most prevalent ways for people who are deaf or hard of hearing to communicate with others.

Sign/Gesture language is a nonverbal communication method which replaces spoken words with hand gestures and movements, body language, and facial expressions (oral communication) [10]. Most of the people communicate using both words and signs, but Deaf and Dumb persons only use signs to convey their emotions and sentiments. In today's modern world, there are numerous types of sign languages available such as ASL (American Sign Language), BSL (British Sign Language), ISL (Indian Sign Language), and so on. People in today's modern world utilize/use their regional sign language to communicate and convey feelings so that others can readily comprehend. A sign language is made up of three key components [3]:

* *Corresponding author.*
E-mail address: 18341A1229@gmrit.edu.in

Finger-Spelling is one of the most important or significant aspects, which implies that there is a symbol for each and every letter of the alphabet. This style of communication is primarily used to spell names and, on occasion, geographical/location names. This is sometimes used to convey words/terms that do not have any particular signs, or to emphasize or explain a certain term/word.

The other/second element is word-based sign vocabulary, which is one of the most often utilized forms of communication among individuals with disabilities. It indicates that in sign language, there is a matching related sign for each word in the vocabulary.

Non-manual elements are the next and the last important aspect of any sign communication. This communication style makes use of facial expressions, lips, body posture, as well as the tongue, among other things.

Two strategies to sign language recognition are possible: vision-based and glove-based. In glove-based technique gloves are worn which consist of sensors. The major disadvantages of these glove-based techniques are, data gloves are relatively costly and these gloves need to be worn continuously which is very difficult hence the modern systems use vision-based approach. It is again divided into two various types such as static approach and dynamic approach. Below figure shows the two types of techniques to language recognition.

(a) **(b)**

Fig. 1 -- Sign Language Recognition Techniques, [a] Glove Based Approach [b] Vision Based Approach

The existing systems are helping society by supporting physically impaired people because it facilitates daily interaction between deaf individuals and the others by integrating multiple sign languages, words, characters and numbers etc.

2. Literature Review

2.1 Real-Time Sign Language Recognition:

The author of this research paper suggested/proposed an automated approach for identifying sign language that uses PCA (Principal Component Analysis) and one-vs.-all SVM (Support Vector Machines) classification. For detecting skin sections color information is used, hand segmentation is accomplished using morphological operations and filters, and feature extraction in hand regions is achieved using PCA, and classification is done with SVM. The system was tuned to function with the five vowels, resulting in a testing accuracy of more than 80% and a frame execution time of 59 milliseconds. This system gives a wide perspective of real-time sign language, which is beneficial to people who are visually handicapped.

2.2 Real-Time Recognition of Indian Sign Language:

Sign languages are made up of hand signals and face emotions. To recognise those signs, OpenCV's skin segmentation technique is utilized/used to detect and track the Regions of Interest (ROI). To train and predict hand gestures, the Fuzzy c-means Clustering which is a machine learning technique is used. Recognizing gestures/signs has been used in many applications, including sign-controlled automation, game control, automated houses, HCI (Human-Computer Interaction), and sign language interpretation, among others. The FCM beats traditional clustering algorithms in terms of efficiency and dependability in many circumstances.

2.3 Conversion of Sign Language into Text:

This research describes two approaches to the recognition of hand gesture/sign: glove/sensor-based and vision-based. A glove-based approach cannot be extensively adopted since data gloves are somewhat expensive. The suggested system employs a vision-based non-

invasive recognition approach. Vision-based recognition can be accomplished using either static or dynamic recognition. This paper demonstrates the use of MATLAB to recognise 26 hand movements in Indian sign language. In this system it includes four modules: pre-processing and hand segmentation, extraction of features, signs identification, and signs to text. Image processing is used to segregate hand motions. Eigenvalues and Eigenvectors are gathered and employed in recognition. To recognize gestures, the LDA (Linear Discriminant Analysis) method was utilized, and the discovered motions were transformed into text and audio format.

2.4 Sign Language Translation:

Sign gestures are captured using the Python OpenCV package through the system's inbuilt camera. The dataset for different signs has been gathered. For each sign, approximately 3000 photos are captured. To analyse and classify visual images, a convolutional neural network is used. CNNs are multilayer perceptron variants that have been regularized. This consists of Convolutional, pooling, flattening, and fully linked layers, as well as activation functions.

2.5Sign Language Recognition for Deaf and Mute:

This system is developed and implemented in such a way that hand movements are identified by recognising and tracking hand gestures with OpenCV's skin segmentation feature. The image frames are changed to maintain uniformity across all films, and OpenCV is used to extract features. The CNN (Convolution Neural Network) technology is used for video classified hand movements and for hand signs training and prediction. In many cases, the CNN outperforms conventional clustering methods in terms of efficiency and dependability. The recognised motions/gestures are matched to the trained models, and the words that correspond to them are predicted. Dual communication has been accomplished in this suggested system, and sensors are not necessary.

2.6 Translation of sign language for Deaf and Dumb people:

The proposed system depicts a kind of supervised machine learning known as Support Vector Machines. This system's primary purpose is to evaluate and transform sign language into text & audio. In this they used k-mean clustering on input images, the methods used in this system are segmentation of the hand region from image by removing background of image, feature extraction of hand region and classification of image by comparing with other dataset images to get the text for related signs. Deaf people's real-time visuals are captured and supplied into the pre-processor. After that, Otsu's algorithm is used to complete the feature extraction process. Following feature extraction, the dataset picture is differentiated with the segmented image, and the relevant result is displayed. SVM is used to classify gestures. Following classification, the relevant sign's text is created, and the voice is generated from text using the MATLAB function.

2.7 Object detection in sports: TensorFlow Object Detection API case study:

Object detection is a broad term that refers to computer vision algorithms for locating and categorizing items. Object detection techniques can be used on both static and dynamic images. In sports, computer vision techniques are already widely used. In this work, the TensorFlow Object Detection API was used to identify football players. The SSD concept with Mobile net was used within the API. In this own data has been taken as the dataset from the football matches. This dataset consists of the images captured from video of the match. It was used to train and evaluate the pretrained model.

The Single Shot Multi box Detector (SSD) is different from R-CNN-based techniques in that it does not require a second stage of classification operation. This technique is used for faster approach of real-time detection applications. It is important to note that "SSD with Mobile Net" is a model with a combination of SSD and Mobile Net. The SSD is used for meta-architecture and Mobile Net for the feature extraction.

The TensorFlow Object Detection API is an "open-source framework built on top of TensorFlow" that intends to make "constructing, training, and deploying object detection models" simple. The TensorFlow Object Detection API accomplishes this by providing the large amount of pre-trained object detection models as well as instructions to the user.

3. Comparison of Different Methods

In sign language recognition the main requirement is dataset for the system. American and British sign language datasets are accessible on internet but Arabic and Indian sign languages are not available on the internet. Because the data was not available online, all of the writers who worked on it created a manual dataset. The data augmentation is done on small datasets to generate image variation, which gradually grows the dataset and improves classification.

There are several processes in a sign language recognition system. The dataset must first be pre-processed. Resizing the image and converting it to grayscale are the first two stages in pre-processing. Pre-processing in another way comprises utilizing colour information to locate skin areas, morphological processes, and median filters to find hand segmentation. In an alternative way, the high-intensity

noises from the video frames are removed during pre-processing. Smoothing or blurring is the first and most important stage in pre-processing, and the most common purpose of blurring is to minimize noise. A convolution process with a low-pass box filter is used to create the blurred picture. The pre-processing procedure includes image capture, segmentation, and morphological filtering.

The next phase is extracting the features. This feature extraction is used to increase and maintain the classifier's accuracy while also reducing its complexity. With the use of contours, features may be retrieved.In other methods for extracting features, it includes Eigenvalues and Eigenvectors, as well as linear discriminant analysis. LDA is a dimensionality reduction technique that identifies a linear combination of characteristics that classifies two or more object or event classes. For categorization, the extracted characteristics are put into a neural network. Another approach is to crop the hand region and eliminate the backdrop using the binary picture of the hand. The original image is then multiplied by the cropped image of the hand.

The weight updating in old traditional technique of neural networks is very difficult because of more amount of the neurons. so, we use the convolutional neural network to overcome the drawback by reducing the input pixel values. When compared to conventional classifiers like KNN and SVM, CNN provides better accuracy. The Inception model is a CNN variant. In this CNN convolutional layer, pooling layer and SoftMax layer are arranged parallel to each other. Instead of one above another, which saves time and money in terms of processing and calculation. Longer temporal dependencies are supported by LSTM, which addresses the removing gradient problem of RNN and provides improved results on lengthy sequences of data.

Table. 1 -- Literature Review

S. No	Paper Name	Algorithm/ Method	Year	Accuracy
1.	Real-Time Sign Language Recognition	PCA (Principal Component Analysis) and one-vs.- all SVM classification	2020	above 80% & execution time of 59 msec per frame.
2.	Real-Time Recognition of Indian Sign Language	Fuzzy c-means Clustering (FCM) algorithm	2019	75 %
3.	Conversion of Sign Language into Text	Linear Discriminant Analysis (LDA) algorithm	2018	79%
4.	Sign Language Translation	CNN (Convolutional Neural Network) algorithm	2020	92%
5.	Sign Language Recognition for Deaf and Mute	Convolutional Neural Network	2021	97%
6.	Translation of Sign Language for Deaf and Dumb people	SVM (support Vector Machine)	2020	95%
7.	Object Detection in sports	TensorFlow Object Detection API Mobile Net SSD	2020	94%

Gradient depression occurs when the depth of the network increases. A residual neural network is used to solve this gradient depression. It is possible to find the least possible path for the image, which we have given as the input by comparing it with the images in the dataset and this reflects the identified class. Class identification is done by the degree of belonging to the cluster centres, fuzzy c-means (FCM) classify the input value to the accurate class. As compared to other clustering methods FCM is more accurate and trustworthy but it takes longer to compute.

4. Conclusion

A survey on sign language translation and recognition is presented in this paper, as well as several strategies for the same have been studied and analyzed. According to table 1, there are numerous techniques present in modern environment to identify sign language.Among all of the research works examined in this paper and table, one aspect that stood out was the quantity and quality of the data set utilized to train the model. It has been demonstrated that as the dataset size increases, the person achieves more accuracy. The quality of the data set is another factor that has a large impact on the model's accuracy, if there is a dataset created by a webcam, a data set with a better resolution the accuracy of the model will increase. The majority of the studies examined are built on CNN and some other methods such as SVM (Support Vector Machines), Region-Based CNN (r-CNN), Matlab with LDA and KNN (K-nearest

Neighbors) etc. In the recognition process, segmentation is important because it divides the skin region from the background of the image, this improves accuracy of the system. Apart from segmentation & classification relies on feature extraction techniques to reduce dimensionality and reduce computational cost. Deep neural networks and TensorFlow Object Detection outperforms classical classifiers such as KNN and SVM in a study of several classification techniques.

REFERENCES

1. C. Amaya and V. Murray, "Real-Time Sign Language Recognition," 2020 IEEE XXVII International Conference on Electronics, Electrical Engineering and Computing (INTERCON), 2020, pp. 1-4, doi: 10.1109/INTERCON50315.2020.9220241.

2. H. Muthu Mariappan and V. Gomathi, "Real-Time Recognition of Indian Sign Language," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), 2019, pp. 1-6, doi: 10.1109/ICCIDS.2019.8862125.

3. Kumar, M. Conversion of Sign Language into Text. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 9 (2018) pp. 7154-7161. Research India Publications. https://www.ripublication.com/ijaer18/ijaerv13n9_90.pdf

4. R. Harini, R. Janani, S. Keerthana, S. Madhubala and S. Venkatasubramanian, Sign Language Translation, 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), pp. 883-886, doi: 10.1109/ICACCS48705.2020.9074370.

5. Dr. M.P. Chitra, Vaishnavi Devi. R, Shalini M, Sriee Sathana. L.B. Sign Language Recognition For Deaf and Mute. International Journal of Recent Technology and Engineering (IJRTE) 4369-4372, Volume-8 Issue-4, April 2021. Available at: https://www.irjet.net/archives/V8/i4/PIT/ICIETET-120.pdf

6. S. Suthagar, K. S. Tamilselvan, P. Balakumar, B. Rajalakshmi, & C. Roshini . Translation of sign language for Deaf and Dumb people. International Journal of Recent Technology and Engineering (IJRTE) 4369-4372, Volume-8 Issue-5, Jan 2020. www.ijrte.org/wp-content/uploads/papers/v8i5/E6555018520.pdf.

7. Mustamo, Pirkko. “Object detection in sports: TensorFlow Object Detection API case study.” (2018). http://jultika.oulu.fi/files/nbnfioulu-201802081173.pdf

8. Divkar, Akshay & Bailkar, Rushikesh & Pawar, Dr. (2021). Gesture Based Real-time Indian Sign Language Interpreter. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 387-394. 10.32628/CSEIT217374.

9. H. Y. Lai and H. J. Lai, "Real-Time Dynamic Hand Gesture Recognition," 2014 International Symposium on Computer, Consumer and Control, 2014, pp. 658-661, doi: 10.1109/IS3C.2014.177.

10. Shruthi, Proddutur, and Dr Jitendra Jaiswal. "A Comparative Analysis On Sign Language Prediction Using Machine Learning Algorithms." (2020). http://www.jetir.org/papers/JETIR2006067.pdf

11. K. Amrutha and P. Prabu, "ML Based Sign Language Recognition System," 2021 International Conference on Innovative Trends in Information Technology (ICITIIT), 2021, pp. 1-6, doi: 10.1109/ICITIIT51526.2021.9399594.

12. M. Xie and X. Ma, "End-to-End Residual Neural Network with Data Augmentation for Sign Language Recognition," 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chengdu, China, 2019, pp. 1629-1633, doi: 10.1109/IAEAC47372.2019.8998073.

13. G. A. Rao, K. Syamala, P. V. V. Kishore and A. S. C. S. Sastry, "Deep convolutional neural networks for sign language recognition," 2018 Conference on Signal Processing And Communication Engineering Systems (SPACES), Vijayawada, 2018, pp. 194-197, doi: 10.1109/SPACES.2018.8316344.

14. Tzutalin (2017) LabelImg. Git code. GitHub repository.

15. A. Das, S. Gawde, K. Suratwala and D. Kalbande, "Sign Language Recognition Using Deep Learning on Custom Processed Static Gesture Images," 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, 2018, pp. 1- 6, doi: 10.1109/ICSCET.2018.8537248.