Proceedings of the 17th INDIACom; INDIACom-2023; IEEE Conference ID: 57626
2023 10th International Conference on "Computing for Sustainable Global Development", 15th - 17th March, 2023
Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi (INDIA)

A Review of Segmentation and Recognition Techniques for Indian Sign Language using Machine Learning and Computer Vision

Subhangi Kumari
Department of Computer Science & Applications
School of Engineering & Technology
Sharda University
Greater Noida, India
subhangisarkar11@gmail.com

Ernest Tarlue
Department of Computer Science & Applications
School of Engineering & Technology
Sharda University
Greater Noida, India
tarluee@gmail.com

Aissatou Diallo
Department of Computer Science & Applications
School of Engineering & Technology
Sharda University
Greater Noida, India
aissatoudiallo181@gmail.com

Megha Chhabra
Department of Computer Science & Engineering
School of Engineering & Technology
Sharda University
Greater Noida, India
megha.chhbr@gmail.com

Gouri Shankar Mishra
Department of Computer Science & Engineering
School of Engineering & Technology
Sharda University
Greater Noida, India
gourisankar.mishra@sharda.ac.in

Mayank Kumar Goyal
Department of Computer Science & Engineering
School of Engineering & Technology
Sharda University
Greater Noida, India
mayankrkgit@gmail.com

Abstract — Indian Sign Language (ISL) is a complex and diverse language used by the deaf and hard-of-hearing community in India. ISL recognition and segmentation have been active areas of research in recent years due to the growing demand for assistive technologies for this population. In this review paper, we provide an overview of the various techniques used for ISL recognition and segmentation using machine learning and computer vision. We begin by discussing the challenges involved in ISL recognition and segmentation, such as the variations in sign language gestures, lighting conditions, and background noise. We then review the literature on various techniques used for ISL recognition. We analyze the strengths and limitations of each technique, comparing their accuracy, speed, and robustness against various factors. Finally, we present the recent advancements and future directions for ISL recognition and segmentation

Keywords — Sign Language, gestures, hand symbols, Segmentation, Recognition, Machine Learning, Indian Sign Language.

I. Introduction

Facial expressions, body posture, and hand gestures can convey ideas in sign language. Its primary users are deaf people and those who can hear but cannot talk. Also, it enables communication between the general population and those who are deaf or inabilities to hear [1]. It is usually used by deaf family members, relatives, and interpreters. Even the shy can communicate with the general public without an interpreter if a sign language recognition (SLR) system operates. People who are deaf or unable to hear frequently utilize these gestures to express their opinions [2]. Such people have trouble communicating with the general population in public places like banks, hospitals, and post offices. To communicate their point of view to a deaf person or to explain their point of view to a deaf person, deaf persons may require the assistance of a sign language interpreter [3]. In contrast, this approach is highly costly and has no long-term advantages for deaf people. So, we need a device capable of automatically decoding sign language gestures. There are regional differences in the sign languages that are utilised. The most popular languages are American sign language (ASL), British sign language (BSL), Signed English (SE), Sign Supported English (SSE) and Pidgin Sign English (PSE).

A. *Plain Indian Sign Language*

The term "Plains Indian Sign Language" often referred to as "Hand Talk", "Plain Sign Talk", or "First Nation Sign Language" refers to a once-common trading language utilised in several modern plains' nations. Northern Mexico, Central and Western America, and Central Canada. Also, it was utilised by the deaf for daily use and oratorios, speeches, and numerous celebrations. There is no evidence connecting spoken language to plain sign language; some people think it is a hand-encoded language or language.

II. Literature Review

We have covered a variety of relevant work on sign language recognition in this area of the literature review. In this instance, even though we have discussed many methodologies and techniques, we have concentrated not only on Indian sign language. However, we have also reviewed work relating to other sign languages. Each associated piece of art offers a unique suggestion for recognizing sign language. Recent literature on Indian sign language recognition using Feature extraction has highlighted the importance of selecting appropriate features for accurately recognizing sign language gestures [2]. Hand-crafted Features such as Histogram of oriented gradients (HOG), Local binary patterns (LBP) [4], and Edge orientation histogram (EOH) have been widely used for ISL recognition, but their performance can be limited by the complexity and variability of sign language gestures. [5] Deep learning-based feature extraction techniques, such as Convolutional Neural Networks (CNNs) [6] and Autoencoders have shown promising results in improving the accuracy of ISL recognition.

A. Indian Sign Language

Indian sign language (ISL) has syntax similar to other spoken languages. ISL employs a multi-layered method for summarization. Based on its structural layout, ISL employs a variety of word orders. Figure 1 shows the Indian Sign Language alphabet. The time SOV framework is typical of ISL. ISL follows its grammar norm, according to [7]. Later, [8] emphasized numerous ISL grammar patterns. A linguistic examination of ISL demonstrates that the syntax is the same all over India.

B. American Sign Language

American sign language (ASL) [9], a branch of French sign language, is the most common sign language used in the United States and most of Canada's English-speaking provinces. Between 2,50,000 and 500,000 people are thought to utilise ASL. In West and Central Africa, the ASL dialect is also used. Many phoneme components, including facial and hand gestures, create ASL characters. Recent work on American Sign Language understanding is described in the following subsections. Figure 2 shows the American Sign Language alphabet.

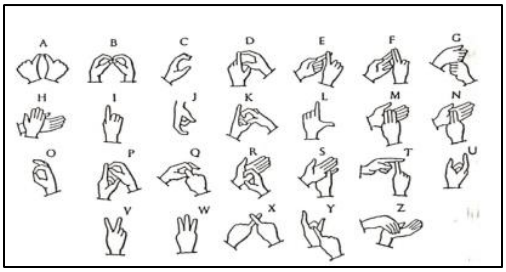

Fig. 1. Indian sign language alphabets

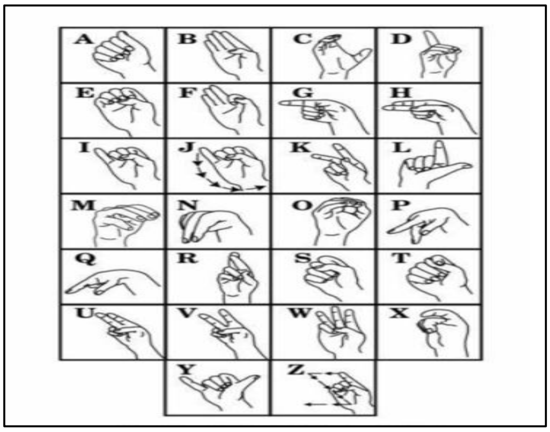

Fig. 2. American Sign Language Alphabets

TABLE I. WORKS RELATED TO THE SIGN LANGUAGE RECOGNITION SYSTEM

S/N	Title/ Year [Ref]	Algorithm/ model used	Dataset	Result	Features	Future scope
1	Deep learning for Sign Language Recognition: Current Techniques, Benchmarks and Open Issues/2021 [5].	CNN, Recurrent Neural Network (RNN) Models , Hidden Markov Models, SVM, Random Forest	Collection of hand gestures in more than 12 sign languages.	Recognized more than 90%; effectiveness above 97%	Should achieve more reliable operation and eliminate some common confusion.	Should achieve more reliable operation and eliminate some common confusion points.
2	Real-Time Vernacular Sign Language Recognition Using MediaPipe and Machine Learning/ 2021 [9].	Hand Landmark models, SVM, KNN, Random Forest, Decision Tree, Naïve Bayes, ANN, MLP	Images of: ASL (alphabet) Indian (Alphabet) Italian (alphabet) ASL (numbers) Turkey (numbers)	SVM=99.29%, KNN=98.8%, Random forest=98.59%, Naïve Bayes=86.7%, ANN=94.79%, MLP=96.48%	Each sign language image dataset is preprocessed to extract features using the MediaPipe framework and trained in Support Vector Machine to classify gestures correctly.	There is a possibility to extend our work by implementing sign language word detection in video using the latest the Mediapipe and the best possible classification algorithms.
3	Multi-Semantic discriminative feature learning for Gesture Recognition using Hybrid Deep Neural Architecture/ 2022 [10].	Multi-class SVM classifier,	Chinese Number Sign Language dataset from the CGD 2011 repository, each consisting of 47 video sequences	Gesture Set 1= 89.63%, 92.313%; Gesture Set 2=77.59%, 90.83%; Gesture Set 1 + Gesture Set 2 = 81.48%, 87.67%	This recognition system works faster than other techniques based on hand analysis and ones developed using high-level features.	It improves classification accuracy if high-level features like optical flow information and motion gradient information get explored.
4	Sign Language Recognition Using Multiple Kernel Learning: A Case Study of Pakistan Sign Language/ 2021 **[11]**.	Hand Segmentation technique, Kernel Selection Algorithm	Developed Images of Pakistan Sign Language for 46 static alphabets	Fuzzy classifier=86%; Template matching=78.20 % & 84%; SVM=83%; SVM with MKL=91.98%	The proposed method takes an image of a bare hand as input, and in the next step, it is segmented using K-means clustering segmentation.	This paper suggests using other functions, such as the Mest Stable Outer Regions (MSER) function instead of SURF.
5	Bangla Numerical Sign Language Recognition Using Conventional Neural	D-LBP technique, convolution neural network model (CNN)	Self-developed images of ten hand signs (a total of 310 pictures)	The Bengali numerical sign language detection	The proposed model takes about a minute to identify Bengali digit characters from an input	We plan to add all Bengali alphabets and integrate the platform into laptops and IoT

	Networks/ 2021. [12]			achieved an overall 99.8% accuracy. Consisting of Precision, Recall, True Negative, Accuracy	image.	media, especially smartphones.
6	A hybrid approach for Bangla Sign Language Recognition/ 2023 [13].	Feature extraction, genetic algorithm, artificial neural network	24 Indian Sign Language alphabets, 10 patterns each, a total of 240 camera images.	The improved recognition rate of 97%	Eigenvalue-weighted Euclidean distance between eigenvectors can be used as a classification method.	Tends to deal with dynamic gestures in future

III. Segmentation and Recognition System

For the overall system, image acquisition, picture pre-processing, separation, feature extraction, and classification are the five steps that comprise the process of Vision-based SLR. The initial step in understanding SL is image acquisition, which can be carried out using either independently developed or openly accessible public datasets. It improves the quality of the image and the extra noise is eliminated in the second preprocessing phase. Preprocessing is followed by segmenting and extracting the region of interest from the overall appearance. For recognition, the fourth stage, feature extraction, converts selected areas of the image we input into feature vectors. The classification phase of vision-based SLR entails identifying the target sign by contrasting its properties with those already noted in the database [14].

A. Image Acquisition Devices

Researchers have used several picture-capturing tools to classify photos. This technology includes a camera or webcam, a data glove, a Kinect, and jump controls. Contrary to data glove-based systems, a camera or webcam is the instrument that most researchers employ since it offers better and more natural interaction with no need for extra equipment. Data gloves have shown to be more accurate in data collection, despite being relatively pricey and cumbersome for users. Many people use Kinect because it works well. It simultaneously delivers a color video stream and a depth video feed. It restores 3D hand motion trajectories and easily detaches the backdrop from the actual sign image [5]. The method for acquiring an image using image acquisition equipment is shown in Figure 3. 43,750 depth pictures—1,250 images for each of the 35 hand gestures—make up our dataset. His five separate subjects provided them. In addition to numbers (0–9) and letters (A–Z), gestures also include the number 2, which is the same as the letter v. The picture is 320x240 in quality and is in grayscale. Figure 3 shows the flowchart of image acquisition.

B. Image Preprocessing Techniques

Preprocessing procedures enhance the quality of an image and eliminate unwanted noise. Alternatively, you might combine a number of these techniques with the original image to accomplish this goal, such as resizing, color conversion, and noise reduction. With the proper preprocessing methods, the process' output can significantly affect accuracy. The two main subcategories of image preprocessing methods are enhancement and restoration. Several image improvement approaches include exponential conversion, contrasting restricting responsive histogram, adaptive histogram equalization, and histogram equalization (CLAHE) [15]. Several filters, including wiener, median, mean, Gaussian, and adaptive, are used in picture restoration [16].

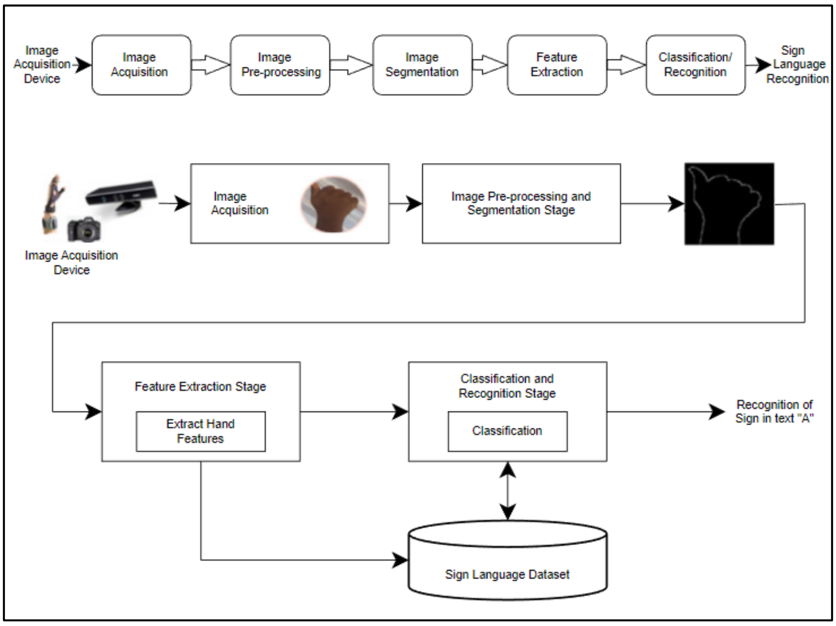

Fig. 3. Image Acquisition of our proposed system

C. *Image Segmentation and Recognition Techniques*

Segmenting an image involves separating it into separate areas or segments [17] [18]. The pictures are separated to extract the area of interest. The two main methods of Segmentation are Contextual segmentation and Non-Contextual segmentation. Contextual Segmentation uses linkages between image components such as edges, same intensities, and near geographic proximity. Instead of considering the spatial relationships between image components, a non-contextual segmentation group bases its pixel value on the value of global attributes. Methods for segmenting images using artificial neural networks include edge detection, thresholding, area, and clustering.

Table 1 summarizes the comparative review along with the future scope of segmentation and recognition techniques along with performance outcomes on the respective dataset used.

IV. Future Scope

There are several potential future directions for research on ISL recognition and segmentation using machine learning and computer vision. Some of these directions are:

Large-scale datasets: The current datasets used for ISL recognition and segmentation are limited in size and diversity. Future research can focus on collecting large-scale datasets with diverse signers, backgrounds, and lighting conditions. This can help improve the robustness and generalizability of ISL recognition and segmentation models.

Multimodal approaches: Sign language recognition can benefit from integrating multiple modalities, such as vision, speech, and gesture. Future research can explore the use of multimodal approaches to improve the accuracy and speed of ISL recognition and segmentation.

User-centric design: Assistive technologies for the deaf and hard-of-hearing community should be designed with user needs and preferences in mind. Future research can focus on developing user-centric design principles for ISL recognition and segmentation systems.

Integration with natural language processing: Sign language recognition can be integrated with natural language processing to enable sign language-based communication with non-signers. Future research can explore the use of natural language processing techniques to improve the accuracy and fluency of sign language-based communication.

In conclusion, future research on ISL recognition and segmentation can focus on developing large-scale datasets, multimodal approaches, online recognition techniques, user-centric design principles, and integration with natural language processing. These directions can help improve the accuracy, speed, and usability of ISL recognition and segmentation systems and benefit the deaf and hard-of-hearing communities in India.

V. Conclusion

Reviewing the many linked works, we have discussed their suggestions for approaches, techniques, models, and other research investigations. We conclude that ISL recognition and segmentation are challenging tasks, and no single technique can address all the complexities of the problem. However, with the rapid advancements in machine learning and computer vision, we are optimistic about the future of ISL recognition and segmentation and their potential to improve the lives of the deaf and hard-of-hearing community in India.

VI. References

[1] V. J. P. P. N. P. M. H. Priyank Mistry, "Indian Sign Language Recognition using Deep Learning," International Confrence on Artificial Intelligence and Machine Vision (AIMV), 2021.

[2] P. C. B. Vaishali Kulkarni, "Indian Sign Langugae Translator Using Gesture Recognition Algorithm," IEEE International Conference on Computer Graphics, Vision and Information Security, 2015.

[3] M. K. T. A Agarwal, "Indian Sign Language (ISL) Biometrics for Hearing and Speech Impaired persons: review and recommendation," International Conference on Contemporary Computing, 2013.

[4] A. Kumar, A. Shukla, and S. S. Kumar, "A Review on Indian Sign Language Recognition," in Proceedings of the 5th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2019, pp. 1187-1191.

[5] T. K. R. S. Muhammad Al-Qurishi, "Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks and Open Issues," IEEE Access, vol. 9, 2021.

[6] A. J. A. C. S. S. K. A. G. Vanita Jain, "American Sign Language Reconition using Support Vector Machine and Convolutional Neural Network," International Journal of Information Technology, 2021.

[7] G. R. Sinha, "Indian sign language (ISL) biometrics for hearing and speech impaired persons: review and recommendation," Int. J. Inf. Technol., vol. 9, pp. 425-430, 2017. doi: 10.1007/s41870-017-0049-0.

[8] B. Hisham and A. Hamouda, "Arabic sign language recognition using Ada-Boosting based on a leap motion controller," Int. J. Inf. Technol., vol. 13, pp. 1221-1234, 2021. doi: 10.1007/s41870-020-00518-5.

[9] A. Halder and A. Tayade, "Real-time vernacular sign language recognition using mediapipe and machine learning," Journal homepage: www.ijrpr.com, ISSN 2582-7421, 2021.

[10] R. E. V. S. L. P. A. G. M. M. B. K. K. L. A. G. A. A. E. Rajalakshmi, "Multi-Semantic Discriminative Feature Learning for Sign Gesture Recognition Using Hybrid Deep Neural Architecture," IEEE Access, 2022.

[11] M. S. S. W. A. A. M. R. O. M. D. S. A. Farman Shah, "Sign Language Recognition using Multiple Kernel Learning: A Case Study of Pakistan Sign Language," IEEE Access, vol. 9, 2021.

[12] S. C. M. M. B. M. K. N. S. S. S. S. F. M. Javed Mehedi Shamrat, "Bangla Numerical Sign Language Recognition using Convolutional Neural Networks," Indonesian Journal of Electrical Engineering and Computer Science, vol. 23, 2021.

[13] M. S. I. N. H. N. N. S. H. W. Sunanda Das, "A hybrid approach for Bangla sign language recognition using deep transfer learning model with random forest classifier," Elsevier , vol. 213, 2023.

[14] A. K. Sahoo, "Indian Sign Language Recognition Using Machine Learning Techniques," vol. 397, 2021.

[15] M. Chhabra, M. K. Shukla, and K. K. Ravulakollu, "Intelligent Optimization of Latent Fingerprint Image Segmentation using Stacked Convolutional Autoencoder," Int. J. Perform. Eng., vol. 17, no. 4, pp. 375-382, 2021.

[16] V. Jain, A. Jain, A. Chauhan, et al., "American Sign Language recognition using Support Vector Machine and Convolutional Neural Network," Int. J. Inf. Technol., vol. 13, pp. 1193-1200, 2021. doi: 10.1007/s41870-021-00617-x.

[17] O. O. B. M. A. A. I. A. Adeyanju, "Machine Learning Methods For Sign Langugae Recognition: A critical Review and Analysis," Elsevier, vol. 12, 2021.

[18] A. Kumar and R. Kumar, "A novel approach for ISL alphabet recognition using Extreme Learning Machine," Int. J. Inf. Technol., vol. 13, pp. 349-357, 2021. doi: 10.1007/s41870-020-00525-6.