A Survey of Sign Language Recognition Systems

Vaishnavi Jadhav[1], Priyal Agarwal[2], Dhruvisha Mondhe[3], Rutuja Patil[4], C. S. Lifna[5]

Department of Computer Engineering, VES Institute of Technology, Mumbai, India

E-mail: [1]2019vaishnavi.jadhav@ves.ac.in, [2]2019priyal.agarwal@ves.ac.in, [3]2019dhruvisha.mondhe@ves.ac.in, [4]2019rutuja.patil@ves.ac.in, [5]lifna.cs@ves.ac.in

Abstract

Technology is now employed to solve practically all humanity's problems. The vast communication gap between the common people and the deaf community is one of the society's fundamental challenges. Computer-based Sign Language Recognition (SLR) systems that translate gestures into text/speech are now being developed by researchers all over the world. This paper attempts to summarize the current state-of-the-art methodologies used by various researchers in the field of SLR systems, right from collecting the data, image processing and using it to translate the gestures. Moreover, the steps followed to develop a SLR system are discussed, and some of the existing SLR applications are reviewed.

Keywords: Machine learning, hand gestures, image processing, mediapipe, long short-term memory, convolutional neural network, real-time recognition, sign language recognition.

1. Introduction

The World Health Organization (WHO) estimates that India's deaf and hearing impaired population is around 63 million. As a result, there is a considerable communication gap between those with disabilities and the wider public. The use of sign language is a significant part of intercommunal communication. However, the main issue is understanding and responding appropriately, which may create a barrier because not all deaf and hard of hearing, as well as normal people, are familiar with sign language. For example, only around 5,00,000 or roughly 1% of 48 million individuals in the United States with hearing loss utilizes sign language. Hearing loss, in fact, is a spectrum, with different forms of hearing loss and communication options. Some deaf people utilize hearing aids, while for others, sound amplification is inefficient or unpleasant. They may use sign language as their primary

way of communication. Sign language is regarded as a separate language from other spoken languages [1].

Hand motions are the foundation of sign language. There are around 300 sign languages in the world, so there is no universal sign language, although American Sign Language (ASL) is the most widely used. In comparison, Indian Sign Language (ISL) has a poor level of recognition. Despite the fact that deaf schools in India do not use it, professionals in India claim that there is a chronic lack of recognition and use of ISL. Various researchers, scientists, and organizations worldwide have already achieved significant progress in computer recognition of sign languages in order to overcome this problem and reduce the communication gap.

Sign language recognition systems are systems, and platforms, that capture hand/face motions, recognize them, and translate them into spoken language. Different approaches and algorithms are used to build and propose various SLR systems. The general flow followed by researchers includes 4 major steps. The steps are:

- Data accusation - Converting real-world signals to the digital domain for display, storage, and analysis is Data accusation.

- Hand segmentation - Hand segmentation is a texture that segments out the portion of the camera feed where the hand is.

- Feature Extraction - Process of transforming raw data into numerical features that can be processed while preserving the information in the original dataset is done by feature extraction.

- Image Classification - Assigning a label or class to an entire image according to their shape, size, etc. is image classification.

These are some major steps included in the collection of the dataset for the models to be built. These steps are discussed in detail in the further parts of the paper. This paper is a review of several studies released by various researchers around the world.

2. Literature Survey

Table 1 lists the most common sign languages and the datasets that are available for those in this domain. Table 2 shows some of the most widely used sign languages translation

software. In the development of sign language recognition systems, two types of databases are used: Standard databases and databases created by them. Many academics wish to train the system using their own datasets. In addition, numerous standard databases for popular sign languages are available. The general classification of the SLR system is shown in Figure 1. Creating the Sign Language Recognition system is a process that takes place in stages. Many methods and algorithms are applied to create an effective and accurate SLR system.
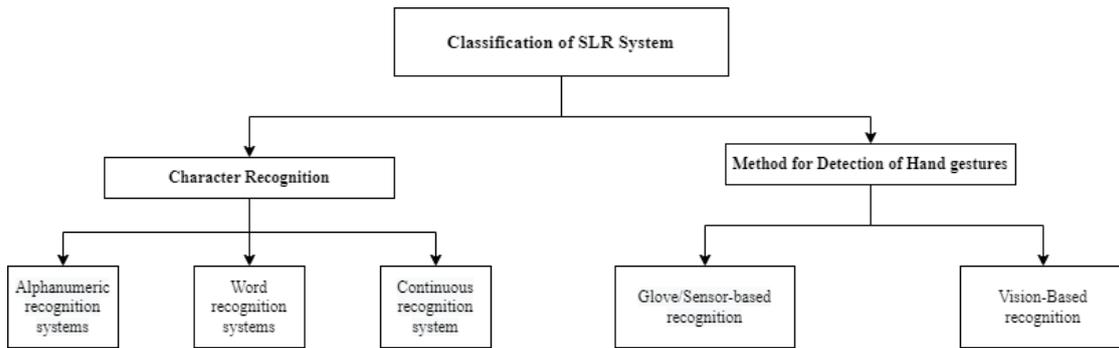

**Figure 1.** General classification of SLR systems

Table 1. Popular sign language methods and datasets [1, 11]

S. No	Sign Language	Abbreviation	Available Standard Datasets
1	American Sign Language	ASL	1. ASL Lexicon Video Dataset 2. Life print Fingerspell Library 3. E-NTERFACE 4. RWTH-BOSTON-104 5. ASL-Kaggle
2	Indian Sign Language	ISL	RWTH-BOSTON-104
3	Brazilian Sign Language	Libras	UCI- Libras Movement Data set
4	Chinese Sign Language	CSL	CAS-PEAL
5	Persian Sign Language	PSL	PETS 2002
6	Spanish Sign Language	LSE	ARAAC
7	Irish Sign Language	ISL	ATIS Corpus
8	Dutch Sign Language	NGT	Corpus NGT
9	German Sign Language	DGS	1. RWTH German Finger-spelling Database 2. RWTH-PHOENIX
10	Australian Sign Language	Auslan	UCI Australian Sign Language

Table 2. Summary of the existing SLR applications

App Name	Language	Recognize				Input	Output	Teaching App	Translation App
		Alphabets	Numbers	Words	Sentences				
Mimix 3D	Signing Exact English (SEE)	✔	✔	✔	✔	Text/ Speech	Gestures	–	✔
Pro-Deaf Translator	ASL, Brazilian, Portuguese	✔	✔	✔	✔	Text	Gestures	✔	✔
The ASL app	ASL	–	–	–	–	–	–	–	✔
Sanket	ISL	–	–	✔	✔	Text	Gestures	–	✔
Indian Sign Language Translator	ISL	✔	✔	✔	✔	Speech	Sign language	–	✔
Sign Language ASL- Pocket Sign	ASL	–	–	–	–	–	–	✔	–
Deaf & Mute Communication	English	✔	✔	✔	✔	Text	Speech	–	✔
Spread Sign	ASL, British, French, German, etc.	–	–	–	–	–	–	✔	–

The general steps followed for the development of the SLR system are shown in Figure 2.
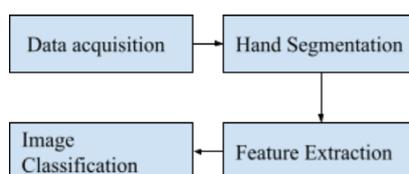
Figure 2. Steps for developing the SLR system

2.1 Data/Image Acquisition

In any image processing system, the first stage is image acquisition. The goal of any image capture is to convert an optical image (real-world data) into a numerical data array that can be edited later, on a computer. In the topic of SLR, there has been a lot of research done, and scholars have created their own standard databases. Researchers utilize various technologies to capture hand motions when creating datasets. The information from the user can be obtained by a variety of means, including visual-based device cameras/web cameras, wearable devices, such as sensor-gloves, cube-gloves, or specific devices such as Microsoft

Kinect. The data acquisition procedure is described as "technical work with photographs" and is briefly discussed in [18].

2.2 Hand Segmentation

The challenge of gesture segmentation is the first step toward visual gesture recognition, which involves detecting, analyzing, and recognizing motions from real-world sequences of images. In SLR systems, hand and gesture detection is critical because the outcome of all other processes is dependent on it. Because there are several different sign tones and segment placements, image segmentation entails training on skin segmentation datasets [2].

The methods [2],[3],[4],[10] consist of the following steps:

- Input image is preprocessed, and the region is computed.
- The image's attributes are preprocessed.
- Transliteration, a method of converting an altered image to text is performed.

2.3 Feature Extraction

The input images contain a lot of irrelevant information that doesn't help with sign language identification, which requires a lot of real-time processing. Feature extraction is the process of converting raw data into numerical features that may be processed while maintaining the information in the original dataset. It yields better results than simply applying machine learning to raw data. Table 3 summarizes the feature extraction approaches as well as the algorithms employed by researchers, along with their outcomes.

2.4 Image Classification

Image categorization is the process of extracting classes of information from a multiband raster image. Thematic maps can be created using the raster created by image categorization. It's crucial in SLR systems since it uses feature-extracted images as input and detects correct gestures. The training set of observational data is usually used to classify the data.

Researchers do not use this as their only means of segmentation. Many scholars utilize new technologies such as MediaPipe as well [8-10], [17]. MediaPipe is used to pre-process images in order to obtain multi-hand landmarks. MediaPipe is a framework that allows developers to create multi-modal applications with a skeleton of nodes, edges, and

landmarks to track key places on the body. All coordinate points are normalized in three dimensions. Developers use Tensorflow for developing models. Pipelines in MediaPipe are made up of nodes in a graph. Only hand movements make up 21 of the crucial aspects. Mediapipe is becoming more popular for SLR development due to the convenience of recording the critical points of gestures.

3. Discussion on the Existing Systems

There are a lot of problems due to which a system fails and there is a need for new systems that overcome these problems. The major issue found in the existing systems is that, most of the applications are sign language teaching apps and not sign language translation apps which can help in the translation of the gestures into the language understood by everyone. While some applications can be used for the conversion to sign language and not vice-versa. The available applications use ASL and not the native language. Also, native languages are not given any preferences over ASL i.e., there are few datasets available for native languages. Apart from this, there is no real-time sign language translator that can be used anytime anywhere. Some existing systems don't recognize the dynamic gestures and also sometimes there is false recognition of the gesture which leads to false communication or wrong learning of the particular language.

Table 3. Summary of the state-of-the-art techniques

Paper	Language	Dataset Used	Recognise	SLR Classification	Device used of Data accusation	Feature Extraction	Image Classification		Output
							Algorithm	Accuracy	
[3]	ISL	Self-made dataset	Alphabets & Numbers	Real-time	Camera	KAZE	K-Nearest Neighbor	97%	Text
							Support Vector Machine	98%	
							Naive Bayes Classifiers	93%	
[4]	ASL	Available ASL dataset	Alphabets A-Z, Del, Space	Real-time	Camera	Canny edge detection	CNN	96%	Text
[5]	ISL	Self-made dataset	A-Z	Real-time	Camera	Euclidean Distance	Fisher's linear discriminant		Text
[6]	SL	Self-made dataset	-	Real-time	Camera	-	-	-	Text
[7]	ASL	-	A-Z	Real-time	Camera	Handshape	Convolutional Neural Network	-	Text & audio

[8]	ASL	-	Alphabets & numbers	Real-time	Camera	-	Support Vector Machine	99%	Text
[8]	ISL	-	Alphabets	Real-time	Camera	-	Support Vector Machine	99%	Text
[9]	ASL	Self-made.	Alphabets, Words	Real-time, continuous	Right hand CYBE RGLOVE sensory glove	Flock of Birds® motion tracker	Light Gradient Boosting	95%	Text
[13]	ASL	Datasets for American WSLR	Words	Real-time	Camera	-	Multi-stream Neural Networks	-	Text
[12]	ASL	Self-made	ASL characters	Real-time	Image	-	CNN	97.3%	Text
[14]	ISL	Self-made dataset	11 alphabets & 23 words	static	Camera	-	KNN	95.95%	Text
							SVM	97.90%	
							ANN	98%	
[19]	ISL	Self-made dataset	26 ISL gestures	-	Camera (White bg)	Morphologi cal filtering	-	-	Text, Speech

4. Conclusion

There are more than 300 different sign languages, one for each country. The development of sign language recognition systems is undergoing extensive research. However, there are numerous flaws in the creation of a fully operational SLR system. All these flaws have been mentioned in this paper. Some of the conclusions derived through this research are summarized as follows: basic issues like the research being limited due to a lack of standard datasets, the creation of a dependable and ubiquitous database is critical for the advancement of systems, the developed technologies operate in situations that are carefully controlled, such as a plain background, clothing color, light intensity, and so on, the currently available systems are focused on static signals; but the development of dynamic communication systems should be prioritized, the creation of a real-time, speedier, and more accurate system should be prioritized, since few movements for distinct signs are the same, such as the hand motion for the number "3" and the alphabet "W" in ASL, it is difficult for a system to recognize them effectively, the number of sign language translation apps available is restricted, most of the applications are educational apps that use video lectures to teach sign languages, there are some translation apps available; these apps accept text or speech as input and use a 3D model to translate the text into motions. To conclude, there are just a few papers that have achieved success in all factors of SLR accuracy, having a huge dataset, and having

real-time recognition being just a few of the factors. More research is being conducted every day on the development of different technologies that help to bridge the gap between the deaf and the normal population. Future work on development of SLR systems should focus on development of real-time SLR system that would work on dynamic gestures as well. Also, the development of SLR systems for the native language apart from ASL is most needed. Moreover, the accuracy of the systems can be increased that would help to prevent the miscommunication that it could cause.

References

[1]   Sahoo, Ashok K., Gouri Sankar Mishra, and Kiran Kumar Ravulakollu. "Sign language recognition: State of the art." ARPN Journal of Engineering and Applied Sciences 9.2 (2014): 116-134.

[2]   Prof. Radha S. Shirbhate1, Mr. Vedant D. Shinde2, Ms. Sanam A. Metkari3, Ms. Pooja U. Borkar4, Ms. Mayuri A. Khandge5. Sign language Recognition Using Machine Learning Algorithm. International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 (3, march 2020).

[3]   Sharath Kumar "Sign Language Recognition with Convolutional Neural Network."International Research Journal of Engineering and Technology (IRJET) Volume: 07 Issue: 12 | Dec 2020.

[4]   NB, Mahesh Kumar. "Conversion of sign language into text." International Journal of Applied Engineering Research 13.9 (2018): 7154-7161.

[5]   Mayuresh Keni, Shireen Meher, Aniket Marathe, Amruta Chintawar, Shweta Ashtekar. Sign Language Recognition System - INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) ICONET – 2014 (Volume 2 – Issue 04).

[6]   Mekala, Priyanka, et al. "Real-time sign language recognition based on neural network architecture." 2011 IEEE 43rd Southeastern symposium on system theory. IEEE, 2011.

[7]   EE, 2011. [8] Halder, Arpita, and Akshit Tayade. "Real-time vernacular sign language recognition using mediapipe and machine learning." Journal homepage: www. ijrpr. com ISSN 2582 (2021): 7421.

[8]   Duy Khuat, Bach, et al. "Vietnamese sign language detection using Mediapipe." 2021 10th International Conference on Software and Computer Applications. 2021. [10] Adhikary, Subhangi, Anjan Kumar Talukdar, and Kandarpa Kumar Sarma. "A Vision-based System for Recognition of Words used in Indian Sign Language Using

MediaPipe." Sixth International Conference on Image Information Processing (ICIIP). Vol. 6. IEEE, 2021.

[9]   Nadgeri, Sulochana, and Dr Kumar. "Survey of Sign Language Recognition System." Available at SSRN 3262581 (2018).

[10]   Jan Matuszewski, Marcin. Zając.Recognition of alphanumeric characters using artificial neural networks and MSER algorithm.Event: Radioelectronic Systems Conference, Jachranka, Poland. (2019)

[11]   Maruyama, Mizuki, et al. "Word-level sign language recognition with multi-stream neural networks focusing on local regions." arXiv preprint arXiv:2106.15989 (2021).

[12]   Sharma, S., R. Gupta, and A. Kumar. "Continuous sign language recognition using isolated signs data and deep transfer learning." Journal of Ambient Intelligence and Humanized Computing (2021): 1-12.

[13]   I.A. Adeyanju, O.O. Bello, M.A. Adegboye, Machine learning methods for sign language recognition: A critical review and analysis, Intelligent Systems with Applications, Volume 12, 2021, 200056, ISSN 2667-3053.

[14]   Liao, Chung-Ju, Shun-Feng Su, and Ming-Chang Chen. "Vision-based hand gesture recognition system for a dynamic and complicated environment." 2015 IEEE International Conference on Systems, Man, and Cybernetics. IEEE, 2015.

[15]   ] Chaikaew, Anusorn, Kritsana Somkuan, and Thidalak Yuyen. "Thai sign language recognition: an application of deep neural network." 2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering. IEEE, 2021.

[16]   Chiu, Liang-Chi, et al. "Fast SIFT design for real-time visual feature extraction." IEEE Transactions on Image Processing 22.8 (2013): 3158-3167

[17]   Sawant, Shreyashi Narayan, and M. S. Kumbhar. "Real time sign language recognition using pca." 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies. IEEE, 2014.

Author's biography

Vaishnavi Jadhav is currently a student pursuing the Bachelor of Engineering in computer engineering at Vivekanand Education Society's Institute of Technology. She is passionate about knowing and working on different technologies. Along with learning different technologies, she enjoys researching the domain of Machine Learning.

Priyal Agarwal is currently a student studying at Vivekanand Education Society's Institute of Technology majoring in Computer Engineering. She has a keen interest in research on various technologies. Recently, she found interest in Machine Learning, and that's the main reason to do this research survey.

Dhruvisha Mondhe is currently a final year student at Vivekanand Education Society's Institute Of Technology, studying computer engineering. She has considerable experience in machine learning and artificial intelligence along with a keen interest in research. This is her third research paper to date. Apart from machine learning, she has also worked in domains of web development and app development.

Rutuja Patil is currently a 4th year student studying at Vivekanand Education Society's Institute of Technology majoring in Computer Engineering. Throughout her 3 years in college, she did various projects on technologies like java, android development, flutter development, and many more. After participating in different hackathons, she gained an interest in learning different technologies.

C. S. Lifna is an Assistant Professor in the Department of Computer Engineering, VES Institute of Technology, Mumbai, India. She completed her B.Tech in Information Technology from Cochin University of Science and Technology, Kerala and M.E in Information Technology from University of Mumbai, Mumbai. Her area of research includes Big Data Analytics, Social Analytics, Text Analytics, Machine Learning, UX Design, and Blockchain.