IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Review On Sign Language Recognition And Translation Systems

[1]Mr. Prathmesh Patil, [2]Miss. Shreya Pansare, [3]Mr. Mayur Bhagade, [4]Prof. P. B. Dumbre

[1]Student, [2]Student, [3]Student,[4]Assistant Professor

[1]Department of Artificial Intelligence and Data Science Engineering,

[1]Jaihind College of Engineering, Kuran Pune

Abstract: The peoples with speech disabilities usually face problems interacting with normal people, and there is need to find a way to make the communication with normal human beings easier. This paper contains a comprehensive review of various systems built to make the communication easier. The systems examined contain a wide range of systems such as gesture detection to the vision based and it holds continuous image capture and recognition which has significant amount of accuracy which utilizes machine learning, convolutional neural network, vision transformer, computer vision and Artificial Intelligence. This paper performs a competitive study on these systems to have a detailed information about it. This paper finds the key differences and the advantages of the systems. There are multiple kinds of datasets are used for the training and testing and this review finds the best datasets. This review will be helpful for the new inventions and for the key findings for future research.

Index Terms - Sign Language, Machine Learning, continuous image capture.

I. INTRODUCTION

Indian sign language recognition (ISL) system helps the speech-disabled people to interact with normal people. A primary communication tool for the hearing-impaired community is sign language [1] [8] [5]. As 70 million deaf people worldwide primarily use sign language [3].

Sign language is a collection of various gesture-generation techniques [6]. Sign language is not just a gesture using fingers and palms; it involves visual cues through the eyes, face, mouth, eyebrows, etc. Additional components, like facial expressions, involve expressing the complex meaning [2] [6]. A sign language recognition system helps physically impaired people to communicate with the rest of the world. People having hearing impairments use gesture-based signs to express their emotions and thoughts [6]. The hard-of-hearing and nonverbal community cannot express their thoughts, ideas, and requirements with general languages such as English, Hindi, Urdu, and Bangla because they cannot talk with their disability [1].

A significant dilemma of Sign Language recognition (SLR) is the lack of publicly available sign language datasets [8]. Multiple people use the same sign for two different meanings at the same time, which makes it a bit confusing for the person to understand.

Demonstrating a sign to look it up may be more natural, but is computationally more challenging due to the rich visual format and linguistic complexities [2]. The numbers don't have some level of confidence because, day by day, some countries immerse with their own sign language. American, British, and Chinese sign languages are the most widely used worldwide [5]. We are reviewing the papers in which the SLR system is build. The main aim behind the review is to find the optimal way to do so or to find the consequences of the building a system in a way. The system can have multiple approaches.

While building the sign language recognition system there is the other main task is to find the Sign Language Translation (SLT) system which could able to find the actual meaning of the signs and covert them to a phrase or line that explains the sign most without making any assumption. The explanation must be clear and must not contain any hidden meaning.

II. MOTIVATION AND BACKGROUND

Sign languages are complex, with large vocabularies and unique phonological rules [2]. Sign language are visual and needed to capture for analysis but the one who is making the sign can use the space around him or even use the things around him to make the best explanation. Stokoe stated in 1960 that signs are composed of hand shape, movement and place of articulation parameters [9]. There are also non-manual components such as mouth patterns [9]. The Motivation for review is that Word-Level American Sign Language had a approximately 90% but while performing on isolated image it has 63% of accuracy. But at the same time American Sign Language Lexicon Video Dataset holds the least amount of accuracy in many systems with approximately accuracy of 34-35%. which describes the system needs a better approach to deal with signs in the form of video.

III. METHODOLOGY

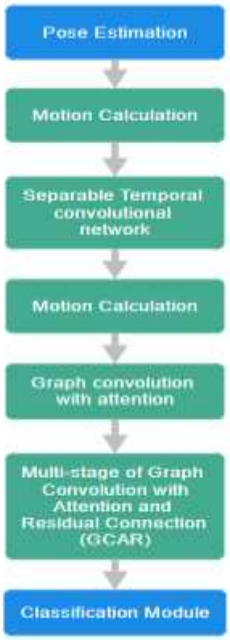
To build the Natural Language Translation System contains multiple complex Task. It also consist of different approaches that have different level of effectiveness. Existing datasets use varied collection and labelling techniques, impacting quality and size [2] [11]. Data scraped from the internet may capture more users or settings, but varied contributor fluency and difficulty identifying and segmenting signing in videos impact quality [2] [13]. Teaching videos are also often professionally-recorded, with similar limitations to lab data. Moreover, scraping is typically not allowed by content creators or hosting platforms [2]. Such kinds of problems are being faced by the system developers. To overcome such a problem, various methodologies are used.

Approaches used for building System :

1. Graph and Convolutional Network based approach:
   This network works with graph theory to find the solution, and it takes advantage of the structural properties of graphs.[1]
2. Natural Language-Assisted Sign Language Recognition (NLA-SLR) :
   Natural Language-Assisted Sign Language Recognition is an approach to improving Sign Language Recognition (SLR) by integrating natural language processing (NLP) techniques.[3]
3. Vision Transformer Based approach:
   This approach splits video frames or images of hand gestures into small patches and transfers into the feature vectors [5].
4. Media Pipe approach:
   This captures the real-time images of the hand gestures and processes them to provide the meaning of the particular sign.

A. Graph and Convolutional nework based approach :

The graph based approach does include some phases to detect and process the image or video to translate the sign. In this approach, the skeleton key points are represented. In this approach, a two-stream model, utilizing a multistage graph convolution with attention and residual connection (GCAR) to extract spatial-temporal contextual information [2]. In the graph based process the joint skeleton points and joint motion points are captured to identify the sign from the sequences. The image size can range from 28*28 to 64*64 according to papers. The detected RBG sign image is converted to HSV (Hue, Saturation, Value) image [11], or gray-scale image[2] [10] or directly passed as a color image [2].



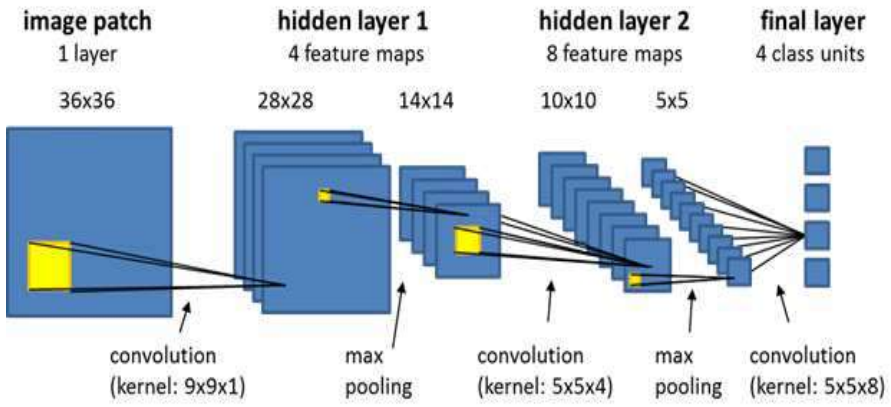


Fig. 2. The basic structure of convolutional layer for approch

Fig. 1. Procedural flow of graph based approach

B. Natural Language-Assisted Sign Language Recognition approach :

Image acquisition is an operation of capturing the images of the hand gesture representing different signs [20]. For skin detection, an adaptive probabilistic model is used. In this model, manually annotated skin and background images are used for creating 32×32×32 RGB colour histograms for both skin and background appearance, and these histograms were normalized and used as probabilistic models of the skin and background [3]. The image size of this approach can range from (256*256), (128*128) , and (64*64) [13] . In this approach various kinds of techniques used such as Histogram Oriented Gradient (HOG) classifiers and there is also some sort of help of SVM in this to learn and recognize the input image [16]. In models it is proposed to model human body keypoints besides RGB videos to enhance the robustness of visual representations [3].
The HOG classifier provides the edges and directions. Images are divided into parrts to calculate the gradient and orientations [3]. The smaller parts are of length (8*8). One of the most important feature or step is feature selection. A word representation learning framework can be adopted to extract gloss features for semantic similarity assessment [13]. There are three main methods to perform feature selection and one of them is Filter method, wrapper method, Embedded method [13]. In this classification and many kind of algorithms are used some systems utilized Support Vector Machine [3][16] and Convolutional Neural Network [16] and Artificial Neural Network [13]. They achieve accuracy ranging from 68\% to 94%.

C. Vision Transformer Based approach :

Vision Transformer was introduced to overcome the challenges occurring in CNNs to capturing long-range dependencies and scaling efficiently. Convolutional neural networks are an artificial neural network.That is combination of layers.In a CNN, an input image passes through convolution layers, where feature detectors create feature maps. Pooling layers then reduce the size and number of parameters, retaining essential features.softmax layer classifies the input based on predefined categories, such as hand gestures[10]. While trasferring image it is converted into patches or small chanks. Those images are directly passed to a linear model and then embedding process is been applied on it and the class token is allocated to each embedded image. They also contain positional patches which increases one by one. Then those patches are provided as input to encoders. CNNs are depends on convolution layers to process images. While Vision Transformer uses the transformer architectures with self-attention mechanisms. The transformer involves encoder and a decoder. Both involve self-attention and feed-forward mechanism. The term of a transformer is use in computer vision is known as a vision transformer This approach gives two important benefit first is Self-attention mechanism , Where model analyzes all input elements simultaneously, allowing it to understand relationships across the entire input context, Second is ability to train on large tasks. Process of Vision Transformer start with converting image into patches by according model design.Then this patches are transferred to the projection layer.

D. Media pipe approach :

The first step is to collect data. The images undergo a series of processing operations whereby the backgrounds are detected and eliminated using the colour extraction algorithm [17]. If while capturing the image the detector isn't able to or fails to detect the hand, then the null entries are added to the matrix and it is removed from the dataset. Segmentation is then performed to detect the region of the skin tone. For the purpose of feature extraction the (64*64) sized image is used [13][17]. Media Pipe is a framework that enables developers for building multi-modal (video, audio, any times series data) cross-platform applied ML pipelines. Media Pipe has a large collection of human body detection and tracking models which are trained on a massive and most diverse dataset of Google [15]. In this multiple processes are involved such as Rotations, Flipping, and scaling.

The algorithms used in this approach are Support Vector Machine (SVM) [15]. SVM is effective in high dimensional spaces. ANd even Artificial Neural Network is employed for the detection of signs. They have varied accuracy, where SVM has the heighest accuracy with the score of 99.07\% and ANN with 97\% of acccuracy. This media pipe approach holds the ability to process multiple photos and videos at a same time which results in a better and optimized results to a sign recognition.

IV. COMPARATIVE RESULT :

All the systems are being reviewed on the basis of how much accuracy they have and how well they can predict the meanings of the signs. There are different approaches used and in them there are different algorithms possesses different accuracy which has the wider range. The Table 1 has summarized algorithms and key findings from the approach and it becomes easy to guess the better approach. The new technologies like media pipe and vision transformer gain more accuracy whereas traditional approaches fail to gain maximum accuracy. There are different kinds of Human Computer Interface used to interact with the system and they are mostly user friendly which ease out interaction for user.

The result can explain us that sign language recognition is a complex task and required sophisticated ways to solve such a problems. The sysem might be complex in nature and might align with time complexity which is the one of the important task to perform in such a systems. The real time prediction is must in these cases which can be main reason for complexity.

Approach	Algorithms Used	Key Features	Accuracy
Graph and Convolutional Network Based	• Graph Convolutional Network (GCN). • Residual Connection (GCAR) • CNN layers for feature extraction. • Softmax for classification	• Extracts spatial-temporal information using skeleton key points • Converts RGB images to HSV/Gray-scale	67%-90%
Natural Language-Assisted Sign Language Recognition (NLA-SLR)	• Support Vector Machine (SVM) • Histogram of Oriented Gradients(HOG) Classifier • CNN and ANN for classification.	• Keypoints-based human body model • Utilizes NLP to enhance recognition	68%-94%
Vision Transformer Based	• Vision Transformers (ViT) • Self-attention Mechanism • CNN for feature extraction	• Splits images into patches • Focus on long-range dependencies in gestures	74%-99%
MediaPipe Approach	• Support Vector Machine (SVM) • Artificial Neural Networks (ANN)	• Real-time processing of hand gestures • Cross-platform applied ML	SVM:99.07 %, ANN:97%

	• Image segmentation and skin-tone detection	pipeline	

V. CHALLENGES WITH SIGN LANGUAGE RECOGNITION AND TRANSLATION SYSTEM :

1) Partial Occlusion and Redundant Backgrounds:
Existing model is depending on image pixel frequently with partial occlusion means situation where part of hand gesture is hidden the camera's view making itdifficult for recognition systems to fully capture it.Redundant background information can complicate for processing of images because it makes it harder for algorithms to focus on the actual signs being performed.This can lead to errors in gesture recognition, it reduces overall accuracy of the system.[1].
2) Gesture Variance:
The gesture variance can be explained as the variance occurred by person or skills of a person or theunderstanding of the person. The every person understands some sign quite differently which could lead to some ambiguity. A person could have different facial expressions which also lead to ambiguity.[1]
3) Limited Generalization:
That models trained to recognize signs may not work well with new signs or new datasets, different signing styles, or changes in the environment, such as background. This makes it difficult for systems to perform in real-life situations.[1]
4) Non-connected skeleton point:
Non-connected skeleton are body joints that do not have direct physical connections, such as the hands and face or the left and right hands, that are not directly connected by the body's structure but are important for recognizing complex sign language gestures that involve coordinated movements across the body.[1]
5) Diverse Sign Languages.
Sign languages vary according regions in terms of gestures, body language, and facial expressions, creating complexity when developing a universal system.These languages use space and facial expressions to add meaning, which makes it difficult to recognize signs accurately.[17]
6) Finger spelling:
Finger spelling is used to represent individual letters of the alphabet, enabling users to spell out names or terms that lack specific signs. While this method is valuable,it can be slower and more difficult for fluent signers,as it requires both the signer to accurately spell words and the viewer to keep up with the spelling.[15]
7) Standard signs:
It represent entire concepts or objects, facilitating smoother and quicker communication. The reliance on finger spelling can disrupt the flow of conversation, especially in situations where time is critical or where quick comprehension is necessary. [15]
8) Lack of Multi-modal Feature Consideration:
The failure of many existing sign language recognition systems to account for various forms of communication that work together in sign language. In sign languages, both manual signs (hand movements) and non-manual features play crucial roles in conveying meaning. When recognition systems focus only on hand movements, they miss out on these important non-manual cues, leading to incomplete or inaccurate interpretations of signs.[8]
9) Real-Time Recognition and Speed :
The proposed method achieves high accuracy, but applying it in real-time sign recognition (such as for video feeds) could still be challenging. Real-time systems require not just accuracy but low latency to provide immediate results [5]

VI. CONCLUSION

The conclusion of this reference paper states that there are new technologies coming that are making the recognition of the signs better and making the predictions much easier.The review compares approximately 4-5 different approaches with examining different datasets. The systems employ various kinds of algorithms like Support Vector Machine, Convolutional Neural Network, Artificial Neural Network and many more algorithms. Different algorithms provide various kinds of accuracy. The accuracy can range from 34% to

99% .The most stable approach seemed is using media pipe and vision transformer based approaches as they are able to process dynamic signs better than other approaches. In conclusion the systems which utilizes the new technologies which are able to process dynamic signs better are performing better.

REFERENCES

[1] S. M. Miah, M. A. M. Hasan, S. Nishimura and J. Shin, "Sign Language Recognition Using Graph and General Deep Neural Network Based on Large Scale Dataset". in IEEE Access, vol. 12, pp. 34553-34569, 2024, doi:10.1109/ACCESS.2024.3372425.

[2] Desai, A., Berger, L., Minakov, F., Milano, N., Singh, C., Pumphrey, K., Ladner, R., Daum´e III, H., Lu, A.X., Caselli, N. and Bragg, D."ASL citizen: a community-sourced dataset for advancing isolated sign language recognition". 2024

[3] R. Zuo, F. Wei and B. Mak, "Natural Language-Assisted Sign Language Recognition," 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 2023, pp. 14890- 14900, doi: 10.1109/CVPR52729.2023.01430.

[4] Zhou, Benjia, Zhigang Chen, Albert Clap´es, Jun Wan, Yanyan Liang,Sergio Escalera, Zhen Lei, and Du Zhang. "Gloss-free sign language translation: Improving from visual-language pretraining." In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp.20871-20881. 2023.

[5] R. Kothadiya, C. M. Bhatt, T. Saba, A. Rehman and S. A.Bahaj, "SIGNFORMER: DeepVision Transformer for Sign Language Recognition," in IEEE Access, vol. 11, pp. 4730-4739, 2023, doi:10.1109/ACCESS.2022.3231130.

[6] D. R. Kothadiya, C. M. Bhatt, A. Rehman, F. S. Alamri and T.Saba, "SignExplainer: An Explainable AI-Enabled Framework for Sign Language Recognition With Ensemble Learning," in IEEE Access, vol.11, pp. 47410-47419, 2023, doi: 10.1109/ACCESS.2023.3274851.

[7] Zheng, Jiangbin, Yile Wang, Cheng Tan, Siyuan Li, Ge Wang, Jun Xia, Yidong Chen, and Stan Z. Li. "Cvt-slr: Contrastive visual-textual transformation for sign language recognition with variational alignment." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 23141-23150. 2023.

[8] Rajalakshmi, E., R. Elakkiya, V. Subramaniyaswamy, L. PrikhodkoAlexey, Grif Mikhail, Maxim Bakaev, Ketan Kotecha, Lubna Abdelkareim Gabralla, and Ajith Abraham. "Multi-semantic discriminative feature learning for sign gesture recognition using hybrid deep neural architecture." IEEE Access 11 (2023): 2226-2238.

[9] De Coster, Mathieu, Dimitar Shterionov, Mieke Van Herreweghe, and Joni Dambre. "Machine translation from signed to spoken languages: State of the art and challenges." Universal Access in the Information Society (2023).

[10] Gangrade, Jayesh, and Jyoti Bharti. "Vision-based hand gesture recognition for Indian sign language using convolution neural network."IETE Journal of Research 69.2 (2023): 723-732.

[11] Abualkishik, Abdallah, Wael Alzyadat, Marwan Al Share, Sara Al-Khaifi, and Mojtaba Nazari. "Intelligent Gesture Recognition System for Deaf People by using CNN and IoT." International Journal of Advances in Soft Computing Its Applications 15, no. 3 (2023).

[12] Alaria, Satish Kumar, Ashish Raj, Vivek Sharma, and Vijay Kumar."Simulation and analysis of hand gesture recognition for indian sign language using CNN." International Journal on Recent and Innovation Trends in Computing and Communication 10, no. 4 (2022): 10-14.

[13] Mohammedali, Atyaf Hekmat, Hawraa H. Abbas, and Haider Ismael Shahadi. "Real-time sign language recognition system." Int J Health Sci 6.S4 (2022): 10384-10407.

[14] A. Singh, A. Wadhawan, M. Rakhra, U. Mittal, A. A. Ahdal and S. K.Jha, "Indian Sign Language Recognition System for Dynamic Signs," 022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2022, pp. 1-6, doi: 10.1109/ICRITO56286.2022.9964940.

[15] 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.

[16] Patil, D. B., and G. D. Nagoshe. "A Real Time Visual-Audio Translator for Disabled People to Communicate Using Human-Computer Interface System." Int. Res. J. Eng. Technol.(IRJET) 8 (2021): 928-934.

[17] Murali, Romala Sri Lakshmi, L. D. Ramayya, and V. Anil Santosh."Sign language recognition system using convolutional neural network and computer vision." International Journal of Engineering Innovations in Advanced Technology ISSN (2020): 2582-1431.

[18] Wadhawan, A., Kumar, P. "Deep learning-based sign language recognition system for static signs." Neural Comput and Applic 32, 7957–7968 (2020).

[19] Tolentino, L. K. S., Juan, R. S., Thio-ac, A. C., Pamahoy, M. A. B.,Forteza, J. R. R.,and Garcia, X. J. O. (2019). "Static sign language recognition using deep learning. International Journal of Machine Learning and Computing", 9(6), 821-827.

[20] Rokade, Y. I.,and Jadav, P. M. (2017). "Indian sign language recognition system. International Journal of engineering and Technology", 9(3), 189-196.