INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 10, October 2025, pp : 1447-1450
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

A REVIEW ON SIGN LANGUAGE RECOGNITION

Rushikesh Takik[1], Prof. Priyanka Yamgekar[2], Akshay Kadu[3], Om Gore[4], Jay Tank[5]
[1,2,3,4,5]Department Of Information Technology, Genba Sopanrao Moze College Of Engineering, Pune, India.
DOI: https://www.doi.org/10.58257/IJPREMS44459

ABSTRACT

The main form of communication for those who have hearing loss is sign language. This language mostly uses non-manual gestures and hand articulations. Recognition of sign language has gained popularity recently. In this paper, we propose a trainable deep learning network that can efficiently capture the spatiotemporal information from a limited number of sign frames for isolated sign language detection. Three networks—the dynamic motion network (DMN), the accumulative motion network (AMN), and the sign recognition network (SRN)—make up our proposed hierarchical sign learning module. In addition, we suggest a method for addressing the variances in the sign samples produced by various signers by extracting essential postures. These crucial postures let the DMN stream acquire the spatiotemporal details relevant to the symptoms. We also provide a cutting-edge method for encapsulating both static and dynamic information about sign motions in a single frame. The main postures of the sign are fused in the forward and backward directions to produce an accumulative video motion frame, preserving the sign's spatial and temporal information. The retrieved features from this frame were combined with the DMN features to be supplied into the SRN for the learning and categorization of signs. This frame was used as input to the AMN stream. The suggested method is effective for recognising solitary sign language, particularly for static signs.

Keywords: Sign Language, Machine Learning, Data Preprocessing, Model Training, AMN, SRN.

1. INTRODUCTION

Globally, hearing loss has become a widespread issue. 23 By 2050, the World Health Organisation projects that 2.5 billion people (or one in four people) would have some degree of hearing loss, with 700 million people needing hearing rehabilitation. The reliance on sign language, the major language of communication for people with varying degrees of hearing difficulties, grows as a result.Sign languages are complete languages with their own grammar and syntax, but they differ from natural languages in terms of their linguistic features . Each sign language 32 has its own dictionary, which is typically smaller than dictionaries for natural languages. As a result, 35 signers utilise a single sign to represent a variety of spoken synonyms, including the phrases "home," "house," and "apartment." Similar to spoken languages, sign languages are diverse; a number of sign languages, including the American sign language 39 (ASL), the Chinese sign language (CSL), and the Arabic sign language (ArSL), are spoken and used around the world . One of the sign languages used in Arabic-speaking nations is ArSL.

Manual and non-manual motions are used together in sign language, which is a descriptive language. The majority of sign language's 63 hand motions used for interpersonal communication are manual motions. These signs are typically 66 accompanied by non-manual motions, which include body 66 stances and expressions. Many sign languages rely heavily on nonmanual gestures to represent 68 emotions and linguistic information that cannot be expressed 69 by manual gestures. For instance, in ArSL, the use of face expressions as adverbs and adjectives to modify manual signs is employed to communicate negations. Additionally, to distinguish between signs that 73 have the same manual gesture, facial expressions are used.

A fundamental part of sign gestures is motion. 77 Signs can be divided into static and dynamic signs based on the motions they include. The majority of sign language letters and numerals are static signs, meaning that they don't move. The forms and positions of the hands and fingers play a major role in these signs .82 Since still photos may effectively capture these kinds of 83 indications, their availability as the majority of the alphabet 84 datasets in the form of photographs is justifiable. Dynamic signals 85, in comparison, entail manual and/or non-manual movements of body parts. 86 The majority of the sign words used in the lexicon of 87 sign languages are represented by these signs.

2. LITERATURE SURVEY

To convey the static and dynamic information of sign gestures in a single frame, suggest an innovative technique. The main postures of the sign are fused in the forward and backward directions to produce an accumulative video motion frame, preserving the sign's spatial and temporal information [1].

The suggested solution will use a deep learning model to automatically recognize hand sign characters and communicate the outcome in Arabic. With a 90% recognition accuracy rate for Arabic hand sign-based letters, this system is guaranteed to be extremely reliable. Advanced hand motions can be used to further increase accuracy.devices for recognition, like Xbox Kinect or Leap Motion. Following the recognition of the Arabic hand sign-based letters, the result is passed into a speech engine, which outputs Arabic audio. [2]

This study uses a variety of traditional machine learning and deep learning models, including one called DeepConvLSTM that combines convolutional and recurrent layers with Long-Short Term Memory cells, to classify This dataset includes 60 signs from American Sign Language recorded with a Leap Motion sensor. To increase the generalization of neural networks, a kinematic model of the left and right forearm, hand, fingers, and thumb is suggested, along with the application of a straightforward data augmentation method. [3]

In this work, we propose a new system based on convolutional neural networks (CNNs) trained on a real dataset to automatically recognize numbers and letters in Arabic Sign Language. To validate our system, we conducted a comparative study demonstrating the effectiveness and robustness of the proposed method compared to traditional approaches based on k-nearest neighbors (KNN) and support vector machines (SVM).[4]

Hand gestures represent a powerful means of communication, serving as a bridge between humans and computers to enable intuitive and natural interaction. Hand Gesture Recognition (HGR) systems aim to support this vision but face several challenges such as gesture irregular ity, illumination variation, background interference, and computational complexity. [5]

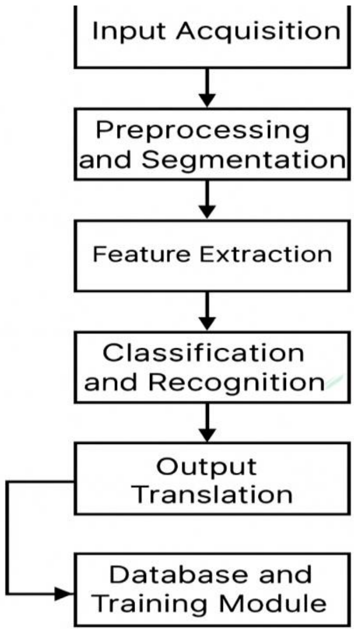
3. METHODOLOGY

The sequential pipeline that makes up the system architecture of a Sign Language Recognition (SLR) model starts with data collection, in which cameras or sensors record hand motions or signals in the form of images or videos. The region of interest is then isolated by passing this input through a preprocessing module that uses hand tracking, noise reduction, and background subtraction. Only the gesture-relevant parts of the frame are processed thanks to the segmentation stage, which increases the precision of the analysis that follows.

The modules for feature extraction and categorization are then activated. Spatial and temporal data are extracted from static or dynamic gestures using methods such as Convolutional Neural Networks (CNN), Artificial Neural Networks (ANN), and Hidden Markov Models (HMM). Following that, these characteristics are categorized into the appropriate sign classes that correspond to letters, words, or phrases. In order to make the identified sign understandable to non-sign language users, it is finally converted into written or audio output. Real-time interaction and conversation between the general public and hearing-impaired people are made possible by the architecture.

4. SYSTEM ARCHITECTURE

The sequential pipeline that makes up the system architecture of a Sign Language Recognition (SLR) model starts with data collection, in which cameras or sensors record hand motions or signals in the form of images or videos. The region of interest is then isolated by passing this input through a preprocessing module that uses hand tracking, noise reduction, and background subtraction. Only the gesture-relevant parts of the frame are processed thanks to the segmentation stage, which increases the precision of the analysis that follows.



INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 10, October 2025, pp : 1447-1450
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

he modules for feature extraction and categorization are then activated. Spatial and temporal data are extracted from static or dynamic gestures using methods such as Convolutional Neural Networks (CNN), Artificial Neural Networks (ANN), and Hidden Markov Models (HMM). Following that, these characteristics are categorized into the appropriate sign classes that correspond to letters, words, or phrases. In order to make the identified sign understandable to non-sign language users, it is finally converted into written or audio output. Real-time interaction and conversation between the general public and hearing-impaired people are made possible by the architecture.

5. COMPARITIVE ANALYSIS

1. Artificial Neural Network (ANN):

ANN is an early and widely used approach in gesture recognition. It simulates the working of biological neurons through interconnected layers. In Sign Language Recognition (SLR), ANN helps identify hand shapes and motions by training on sample datasets. It performs well with **simple gestures and small datasets**, and is suitable for static image-based signs. However, ANN faces issues like **overfitting, slow convergence**, and reduced performance on large, complex datasets. It also lacks the ability to automatically extract deep features, requiring manual preprocessing and feature selection, which limits scalability.

2. Convolutional Neural Network (CNN):

CNN is a deep learning-based architecture that has revolutionized image-based sign recognition. It uses multiple layers for **automatic feature extraction**, such as convolutional, pooling, and fully connected layers. CNNs have demonstrated exceptional accuracy, achieving **up to 99%** in recognizing static and dynamic hand gestures. They are robust against noise, lighting variations, and background clutter. However, CNNs demand **large training datasets, high computational power,** and **GPU-based processing** for real-time performance. Despite this, CNN remains the **most reliable and accurate** method for modern sign language recognition systems.

3. Hidden Markov Model (HMM):

HMM is particularly effective for **dynamic and time-dependent gestures**, making it valuable for recognizing continuous sign sequences. It models gestures as a sequence of states and transitions, capturing temporal variations in motion. The key advantage of HMM is its ability to handle **sequential and time-series data** effectively. However, it suffers when hand segmentation is inaccurate or when gestures overlap, leading to **ambiguity in recognition**. Moreover, training HMMs can be complex and computationally intensive, especially when dealing with high-dimensional data or large gesture vocabularies.

4. K-Nearest Neighbor (K-NN):

K-NN is a simple and non-parametric classifier used for gesture classification based on similarity measures. It classifies gestures by calculating distances (like Euclidean distance) between the new sample and stored training samples. Its advantages include **simplicity, fast implementation, and no training phase.** However, it is **highly sensitive to noise,** has **high memory usage,** and **slows down** as the dataset size increases. K-NN is best suited for small datasets and quick recognition tasks but not ideal for large-scale, real-time applications.

5. Deep Conv LSTM (Hybrid Deep Learning Model):

Deep Conv LSTM **integrates Convolutional Neural Networks (CNN)** for spatial feature extraction and **Long Short-Term Memory (LSTM)** networks for temporal sequence learning. This hybrid model captures both visual and motion dynamics of gestures, making it ideal for **real-time continuous sign language recognition.** It prevents overfitting through data augmentation and handles sequential data efficiently. Despite its advantages, DeepConvLSTM models require **large computational resources** and may show performance variations across different users due to **individual gesture differences**.

6. RESEARCH GAP AND FUTURE SCOPE

A. Current Limitations

Current sign language recognition systems face several limitations that hinder their widespread use. The major drawback is the **absence of standardized and diverse datasets** representing different languages such as ASL, BSL, and ArSL. Variations in **lighting conditions, camera quality, hand occlusion, and background noise** significantly affect model performance. Furthermore, many systems focus primarily on **static gestures**, neglecting continuous and complex sign sequences that are essential for full sentence translation. High **computational requirements** and dependency on specialized hardware limit implementation on mobile or low-cost platforms, reducing accessibility for users in real-time communication environments.

B. Identified Research Gaps

The existing literature reveals several significant research gaps in sign language recognition. There is still a lack of research on **multilingual, real-time sign interpretation systems** that can process continuous gestures across languages. Moreover, **semantic understanding** using **Natural Language Processing (NLP)** remains underexplored, which restricts the ability to convert recognized gestures into contextually meaningful sentences. Another gap lies in the lack of **emotion and context-aware models,** as human communication involves facial expressions and emotional cues in addition to gestures. Additionally, current systems often lack **cross-user adaptability,** meaning they do not perform consistently when trained on one signer and tested on another.

C. Future Enhancement Directions

Future advancements should focus on developing **lightweight, real-time SLR models using transfer learning and edge AI** techniques that can operate efficiently on mobile devices. Building **global standardized datasets** encompassing multiple sign languages would greatly improve training diversity and model generalization. Integrating **speech synthesis, augmented reality (AR), and natural language processing** could make systems more interactive and context-aware. Future research can also explore **multimodal fusion approaches** that combine **visual, skeletal, and sensor data** for enhanced accuracy and robustness. Finally, implementing **emotion and context recognition capabilities** will allow sign language systems to interpret not only gestures but also the **intent and meaning behind communication,** creating a more inclusive and intelligent interaction platform.

7. CONCLUSION

The past ten years have seen a substantial increase in the popularity and interest of sign language recognition among researchers all around the world. Different types of sign acquisition, recognition strategies, target languages, and the quantity of recognised signs have all been proposed for isolated sign language recognition. In this study, we presented three deep learning models for sign language recognition: DMN, AMN, and SRN. The DMN stream successfully learnt the spatiotemporal characteristics of key postures, and we suggested a novel method to extract key postures that successfully handled sign sample variability. In order to benefit the second suggested network, AMN, sign motion was encoded into a single image using the AVM method. With the help of the third network, SRN, which successfully combined features from the DMN and AMN streams, recognition performance was enhanced.

8. REFERENCES

[1] HAMZAH LUQMAN "An Efficient Two-Stream Network for Isolated Sign Language Recognition Using Accumulative Video Motion" IEEE Access 2022.

[2] Hayani, S., Benaddy, M., El Meslouhi, O., & Kardouchi, M. (2021). Arabic Sign Language Recognition with Convolutional Neural Networks. IEEE Access**,** 9, 110214–110225.

[3] Kamruzzaman, M. M. (2022). Arabic Sign Language Recognition and Speech Generation Using Convolutional Neural Networks. Journal of King Saud University – Computer and Information Sciences, 34(6), 3621–3629.

[4] Sharma, R., Singh, P., & Kaur, H. (2022). Sign Language Gesture Recognition Using Deep Learning Approaches. International Journal of Recent Research Aspects**,** 9(2), 14–21.

[5] Hernandez, V., Suzuki, T., & Venture, G. (2022). DeepConvLSTM for Human Activity and Sign Language Recognition. PLOS ONE**,** 17(3), e0265487.

[6] Wu CH, Lin CH (2013), "Depth-based hand gesture recognition for home appliance control". In: Proceedings of IEEE 17th international symposium on consumer electronics, pp 279–280

[7] Liu X, Fujimura K (2004). "Hand gesture recognition using depth data". In: Proceedings of IEEE international conference on automatic face and gesture recognition, pp 529–534

[8] Priyal SP, Bora PK (2010). "A study on static hand gesture recognition using moments". In: Proceedings of international conference on signal processing and communications (SPCOM), pp 1–5

[9] Ren Z, Yuan J, Meng J, Zhang Z (2013). "Robust part-based hand gesture recognition using Kinect sensor". IEEE Trans. Trans Multimedia 15(5):1110–1120

[10] Barkoky A, Charkari NM (2011) Static hand gesture recognition of Persian sign numbers using thinning method. In: Proceedings of international conference on multimedia technology (ICMT), pp 6548–6551

[11] Swapna B, Pravin F, Rajiv VD (2011) Hand gesture recognition system for numbers using thresholding. Commun Comput INF Sci 250:782–786