International Research Journal of Multidisciplinary Sciences (Double-Blind Peer Review Journal)
VOL-2 ISSUE-4 April 2026 PP:38-41 ISSN:3107-930X(ONLINE)

Indian SignLanguage Gesture Recognition System: A Review

[1]RiteshYadav,[2]ShyamolBanerjee
[1]M.TechStudent,Dept.ofCSE,SRCEM,[2]AssistantProfessor,Dept.ofCSE,SRCEM
[1]riteshyadav2100@gmail.com,[2]shyamolgwl@gmail.com

Abstract

Indian Sign Language (ISL) is the primary mode of communication for millions of deaf and hard-of-hearing individuals in India. Despite significant advances in artificialintelligence,computervision,anddeeplearning, communication barriers between hearing and hearing-impairedcommunitiescontinuetoexistduetothelimited adoptionofautomaticsignlanguagerecognitionsystems. Indian Sign Language Gesture Recognition (ISLGR) systems aim to bridge this communication gap by automatically identifying hand gestures, facial expressions, and body movements and translating them into text or speech. Recent developments in machine learning, convolutional neural networks (CNNs), recurrent neural networks (RNNs), transformers, and vision-basedtechnologieshavesignificantlyimprovedthe performance of ISL recognition systems. This review presents an overview of IndianSign Language, discusses theevolutionofgesturerecognitiontechniques,examines existing datasets and recognition methods, highlights recent research trends, and identifies current challenges and future research directions. The review emphasizes that robust, real-time, and cost-effective ISL recognition systems have the potential to improve accessibility in education, healthcare, public services, and daily communication.

Keywords: Indian Sign Language, Gesture Recognition, Computer Vision, Deep Learning, MachineLearning,Human-ComputerInteraction.

I. Introduction

Communication is one of the fundamental requirements of human interaction. However, individuals with hearing and speech impairments often face considerable challenges in communicating with people who are unfamiliar with sign language. According to global health estimates, millions of people suffer from hearing disabilities, making sign language an essential communication medium. In India, Indian Sign Language (ISL) serves as the standardized visual language used by the deaf community. Unlike spokenlanguages,ISLreliesonhandshapes,hand orientation, movement, facial expressions, and body posture to convey meaning.

Traditional communication methods usually requirethepresenceoftrainedinterpreters,which limits accessibility in educational institutions, hospitals, government offices, workplaces, and emergency situations. The rapid advancement of computer vision and artificial intelligence has created new opportunities for developing automated sign language recognition systems capable of translating sign gestures into readable text or synthesized speech. Such systems reduce communication barriers and promote social inclusion.

Early ISL recognition systems primarily depended on sensor-based devices such as data gloves equipped with accelerometers and flex sensors. Although these systems achieved reasonable recognition accuracy, they were expensive, uncomfortable, and impractical for everydayuse.Theemergenceofaffordabledigital cameras and deep learning techniques shifted research toward vision-based gesture recognition systems, which require only a standard camera and sophisticated image-processing algorithms. Modern ISL recognition systems employ convolutional neural networks, recurrent neural networks, transformer-based architectures, and attentionmechanismstorecognizebothstaticand dynamic gestures with high accuracy.

The growing availability of publicly accessible datasets, increased computational power, and advanced graphics processing units have accelerated research in this field. Nevertheless, Indian Sign Language remains comparatively underrepresented compared with American Sign Language and British Sign Language, creating significant opportunities for future research.

II. Indian Sign Language Gesture Recognition

Indian Sign Language is a natural language with itsowngrammar,syntax,andvocabulary.Itisnot avisualrepresentationofspokenHindiorEnglish but an independent linguistic system developed within the Indian deaf community. ISL contains both static gestures, where the hand maintains a fixed position, and dynamic gestures involving continuous hand movement. Many signs also depend on facial expressions, lip movement, and body posture, making recognition a challenging computer vision problem.

Gesture recognition systems generally follow a sequence of stages including image acquisition, preprocessing, hand segmentation, feature extraction, gesture classification, and translation. During image acquisition, images or videos are captured using webcams, smartphone cameras, RGB cameras, or depth sensors. Preprocessing improves image quality by reducing noise, adjusting illumination, and normalizing input dimensions. Hand segmentation separates the hand region from the background using color models,thresholdingtechniques,ordeeplearning segmentation networks.

Feature extraction converts visual information into meaningful numerical representations. Earlier approaches relied on handcrafted features suchas Histogram ofOriented Gradients(HOG), Scale-Invariant Feature Transform (SIFT), Speeded-Up Robust Features (SURF), and Local Binary Patterns (LBP). Modern recognition systems automatically learn hierarchical features using deep neural networks, eliminating the need for manual feature engineering.

Classification algorithms identify the gesture represented by the extracted features. Machine learning classifiers including Support Vector Machines (SVM), Decision Trees, Random Forests, and K-Nearest Neighbors (KNN) were widely used in early systems. Recent studies predominantly employ deep learning architectures suchasCNNsforspatialfeatureextraction,Long Short-Term Memory (LSTM) networks for temporal sequence modeling, and hybrid CNN-LSTM frameworks for continuous sign recognition. More recently, transformer-based visionmodelshavedemonstratedpromising performance by effectively capturing long-range dependencies in gesture sequences.

Machine Learning and Deep Learning Approaches

Machine learning played a significant role in the initial development of ISL recognition systems. Researchers manually extracted geometric and appearance-based features before training classifiers such as Support Vector Machines, Artificial Neural Networks, Naïve Bayes classifiers, and Random Forests. These approaches performed well under controlled environmentsbutsufferedfromreducedaccuracy when background complexity, lighting variation, and user diversity increased.

Deep learning has transformed gesture recognition by enabling automatic feature learningdirectlyfromrawimages.Convolutional Neural Networks have become the most widely adopted architecture for recognizing static ISL alphabets and isolated gestures. CNNs extract multiple levels of spatial features through convolution, pooling, and fully connected layers, resulting in high classification accuracy without handcrafted descriptors.

Dynamic gesture recognition requires understanding temporal information across video frames. Recurrent Neural Networks, particularly Long Short-Term Memory networks, effectively model sequential hand movements and improve recognition performance for continuous sign language. Hybrid CNN-LSTM architectures combine spatial feature extraction with temporal modeling, making them suitable for recognizing complete words and sentences.

Recent research has explored transformer architectures and attention mechanisms, which provide improved context modeling and parallel computation. Vision Transformers (ViTs) have demonstrated superior performance for several computer vision tasks and are gradually being applied to sign language recognition. Transfer learning using pre-trained models such as VGG16, ResNet50, InceptionV3, MobileNet, EfficientNet, and DenseNet has also become popularduetolimitedISLdatasets.Thesemodels reduce training time while achieving competitive recognition accuracy.

TheintegrationofMediaPipehandtracking,pose estimation algorithms, and lightweight deep learning models has enabled the development of real-time mobile applications capable of recognizing ISL gestures with relatively low computational requirements.

III. DatasetsandPerformanceEvaluation

The availability of high-quality datasets plays a crucialroleinthedevelopmentofaccurategesture recognitionsystems.Comparedwithinternational sign languages, publicly available Indian Sign Language datasets remain relatively limited in size and diversity. Existing datasets typically contain alphabets, numbers, isolated words, and selected daily communication gestures captured under controlled laboratory environments. Many datasets suffer from limited participant diversity, background variation, and illumination changes, restricting model generalization.

Researchers commonly evaluate ISL recognition systemsusingmetricssuchasaccuracy,precision, recall,F1-score,specificity,sensitivity,confusion matrices, and Receiver Operating Characteristic (ROC) curves. For continuous gesture recognition, Word Error Rate (WER), Character Error Rate (CER), and sequence prediction accuracy are also employed. Cross-validation techniques and independent testing datasets help estimate the robustness and generalization capability of recognition models.

Recent deep learning models frequently achieve recognitionaccuraciesexceeding95%forisolated gesturedatasets.However,performancegenerally decreases in real-world environments where variations in camera angle, lighting conditions, background clutter, hand occlusion, and signer diversity introduce additional complexity.

IV. Current Challenges and Future Directions

Althoughremarkableprogresshasbeenachieved, several research challenges remain. One of the major limitations is the scarcity of large-scale annotated Indian Sign Language datasets containing diverse users, regional variations, and continuoussentence-levelgestures.Mostexisting datasetsfocusonlyonalphabetsorisolatedwords, limiting practical deployment.

Background complexity and illumination variation continue to affect recognition accuracy in unconstrained environments. Hand occlusion, overlappingfingers,rapidgesturetransitions,and differencesinhandsizefurthercomplicatefeature extraction. Continuous sign language recognition remains substantially more difficult than isolated gestureclassificationbecauseitrequirestemporal segmentation and contextual understanding.

Another challenge is the limited integration of facial expressions and body posture, both of which carry essential grammatical informationin Indian Sign Language. Many existing systems rely solely on hand gestures, leading to incomplete semantic interpretation. Computational complexity also presents challengesfordeploymentonmobiledevicesand embeddedsystemswherememoryandprocessing resources are limited.

Future research is expected to focus on multimodal recognition systems that simultaneously analyze hand gestures, facial expressions, skeletal movements, and contextual information. Transformer-based architectures, graph neural networks, self-supervised learning, federated learning, and multimodal foundation models are promising research directions. Expanding publicly available ISL datasets through collaborative efforts among universities, government organizations, and the deaf communitywillsignificantly accelerateprogress. The integration of ISL recognition with augmented reality, wearable devices, smartphones, and cloud computing will further improve accessibility in education, healthcare, banking, transportation, and public administration.

V. Conclusion

Indian Sign Language Gesture Recognition has emerged as an important research area at the intersection of computer vision, artificial intelligence, and human-computer interaction. The transition from sensor-based systems to vision-based deep learning approaches has substantially improved recognition accuracy while reducing implementation costs. Convolutional neural networks, recurrent neural networks, transformer models, and transfer learning techniques have demonstrated remarkablepotentialforrecognizingbothstatic and dynamic ISL gestures. Nevertheless, the development of practical real-time systems continues to face challenges related to dataset availability,environmentalvariability,continuous gesturerecognition,andcomputationalefficiency. Future research emphasizing multimodal learning, large annotated datasets, explainable artificialintelligence,andlightweightdeployment models is expected to make ISL recognition systems more accurate, scalable, and accessible. Such advancements will contribute significantly to inclusive communication and improve the quality of life for individuals with hearing and speech impairments.

References

1. Koller,O.(2020).QuantitativeSurveyof the State of the Art in Sign Language Recognition.*ComputerVisionandImage Understanding*, 198, 102965.
2. Goodfellow,I.,Bengio,Y.,&Courville, A.(2016).*DeepLearning*.MITPress.
3. Krizhevsky,A.,Sutskever,I.,&Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 25.
4. Vaswani, A., et al. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems*, 30.
5. Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *International Conference on Learning Representations (ICLR).*
6. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.
7. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *MICCAI*, 234–241.
8. Indian Sign Language Research and Training Centre. *Indian Sign Language Dictionary*. Government of India.