Sign Language Recognition System Sign Wave

Abhishek Singh Karki, Anuj Patel
Department of Computer Science
Babu Banarasi Das Institute of Technology and Management
BBD Green City, Faizabad Road, Lucknow, Uttar Pradesh-226028 India
{karkiabhishek2003, anujpatelfzd}@gmail.com

Prof. Richa Verma
Department of Computer Science
Babu Banarasi Das Institute of Technology and Management
BBD Green City, Faizabad Road, Lucknow, Uttar Pradesh-226028 India
Richa.testing5030@bbdnitm.ac.in

Abstract – People with hearing and speech disabilities often struggle to communicate without a translator, as sign language is not universally understood. They rely heavily on hand gestures for non-verbal communication. To address this, the paper proposes a system for automatic recognition of finger spelling in Indian Sign Language. The process begins by capturing the sign as an image input. Skin colour-based segmentation is used to detect hand shape, followed by conversion to a binary image. A Euclidean distance transformation is applied, and row/column projections are performed. For feature extraction, central moments and HU's moments are used, while classification is done using Neural Networks and SVM.

Keywords: Artificial Neural Network, Central Moments, Distance Transformation, Fourier Descriptor, HU's Moments, Indian Sign Language, Projection, Skin Segmentation, SVM.

I. INTRODUCTION

Sign Language Recognition (SLR) is a vital field aiming to bridge the communication gap between hearing-impaired individuals and the general public. Sign languages are rich visual-gestural languages incorporating hand gestures, facial expressions, and body postures, each with its own grammar and syntax. However, their machine recognition is challenging due to variations in signer style, lighting, backgrounds, and regional differences—especially in underrepresented languages like Indian Sign Language (ISL).

While global efforts have primarily focused on American Sign Language (ASL) using specialized hardware (e.g., sensor gloves, depth cameras), such systems are often costly and inaccessible in resource-limited settings. ISL, used by over 2 million people, suffers from lack of standardization and limited technological support.

Recent advances in computer vision and deep learning, especially Convolutional Neural Networks (CNNs), offer the potential for real-time, cost-effective SLR using standard hardware like webcams and smartphones. This project proposes a scalable, robust ISL recognition system capable of interpreting static and dynamic gestures in real time using affordable devices. The aim is to foster inclusivity in education, public services, and assistive technology through practical deployment.

---

2. Scope of the Project

The project focuses on building a **real-time ISL recognition system** that is accurate, user-friendly, and adaptable. Major components include:

- **Gesture Recognition**: Real-time processing of live video input using deep learning to recognize static and dynamic ISL gestures with minimal latency.
- **Scalable Framework**: Modular system architecture that supports vocabulary expansion and regional ISL dialects.
- **User-Centric Design**: Utilizes accessible devices (e.g., webcams, smartphones) with a simple interface suitable for both technical and non-technical users.
- **Accuracy and Performance**: Employs advanced models (e.g., CNNs, RNNs, Transformers) and image processing to ensure

high accuracy (>95%) across varying user conditions, backgrounds, and lighting.

- **Applications**: Includes use in public services, education, workplace communication, emergency scenarios, and human-computer interaction, enhancing accessibility and digital empowerment.

II. LITERATURE REVIEW

1. **Abu Saleh Musa Miah et al. (2024, IEEE)**
This study underscores the transition in Sign Language Recognition (SLR) research from traditional image-processing techniques to advanced graph-based models and deep neural networks (DNNs). With the advent of large-scale datasets, graph-based structures have shown effectiveness in capturing spatial dependencies within gesture data. These deep learning approaches significantly enhance recognition accuracy and robustness, providing a more scalable solution for real-world applications involving complex and dynamic sign gestures.
2. **M. Alaftekin, I. Pacal, and K. Cicek (2024, Springer)**
The authors focus on real-time SLR systems, noting the limitations of earlier methods such as Hidden Markov Models in dynamic environments. With the integration of the YOLO (You Only Look Once) object detection algorithm, the study demonstrates improved speed and real-time performance in gesture recognition. Despite these benefits, challenges such as detecting small hand gestures and managing background interference persist. The paper advocates for hybrid models and optimization techniques to enhance YOLO's performance for SLR.
3. **V. Pavani, Y. Neeharika, and G.S. Ishwarya (2024,IEEEXplore)**
This paper explores a hybrid approach using MediaPipe for pose and landmark estimation combined with Long Short-Term Memory (LSTM) networks for sequential modeling. MediaPipe enhances real-time gesture tracking in uncontrolled environments, while LSTM effectively captures temporal dependencies in dynamic gestures. This combination proves valuable for improving the accuracy and responsiveness of SLR systems in real-world scenarios, though challenges such as environmental noise and lighting variability remain.
4. **Multimedia Tools and Applications (2024, Springer)**
A comprehensive review detailing the evolution of SLR from rule-based and statistical models to CNN and LSTM-based deep learning architectures. The paper emphasizes multimodal approaches, integrating visual and sensor data to improve recognition performance. It also highlights the role of pre-trained models and transfer learning in enhancing generalization across various sign languages. Key challenges identified include real-time responsiveness, background noise, and the need for larger and more diverse datasets.
5. **YSRISAI Praneeth and SU Kiran (2024, ijojournals.com)**
This literature outlines the shift from handcrafted and rule-based gesture recognition techniques to data-driven deep learning models. CNNs have significantly boosted gesture detection accuracy by learning spatial features, while RNNs and LSTMs handle temporal aspects. The paper also points to the emerging importance of 3D pose estimation and multimodal sensor integration to enhance recognition in dynamic, noisy settings. Ongoing challenges include dataset scarcity, gesture complexity, and achieving real-time performance.
6. **K. Balogun (2024, Int. Arab J. Inf. Technol.)**
This review focuses on fingerspelling, a subset of SLR that involves spelling words through hand gestures. It traces the evolution from manual feature extraction to CNN-based models, which improve accuracy and enable real-time recognition. However, the system still faces issues with occlusions, varied hand shapes, and inconsistent lighting. The paper emphasizes the need for optimized CNN architectures and rich datasets to address these practical barriers, especially on edge devices.
7. **A.L.C. Carneiro, D.H.P. Salvadeo, and L.B. Silva(2024,arXiv)**
The authors propose combining deep learning models with low-cost handcrafted descriptors to balance recognition accuracy and computational efficiency. While CNNs offer robust feature

learning, they require significant resources. By integrating them with descriptors capturing shape, motion, and texture, the hybrid system reduces overhead, making real-time applications more feasible on resource-constrained devices. However, dynamic gestures and signer variability continue to pose significant challenges.

8. **A. Deshmukh (2024, shibata.yubetsu)** Targeting Indian Sign Language (ISL), this review outlines how CNNs have enhanced gesture recognition by automatically learning feature hierarchies. ISL's unique gesture patterns introduce challenges in dataset development and model generalization. Despite notable improvements in accuracy, issues such as occlusion, gesture complexity, and limited annotated datasets hinder widespread implementation. The study underscores the potential of CNN-based models to support accessibility in diverse Indian settings.
9. **A.S.M. Miah, M.A.M. Hasan, S. Nishimura, and J. Shin (2024, IEEE Xplore)** This work revisits the integration of graph-based learning and DNNs, emphasizing their role in capturing spatial-temporal dependencies in gesture data. The combination allows for more nuanced recognition of dynamic and complex gestures. Large-scale datasets are highlighted as critical for model training and generalization. The authors also note challenges such as real-time inference, gesture variability, and occlusion management, recommending graph-DNN hybrids as a path forward.
10. **M. Akilan and M.F.A. Lourdes (2024, JARTMS)**
The study investigates hybrid models that combine CNNs with traditional machine learning classifiers like SVM and Random Forest. These systems enhance classification robustness, particularly in scenarios with limited or imbalanced training data. While CNNs automatically extract features, the use of traditional classifiers helps improve model interpretability and generalization. Key limitations include real-time efficiency and scalability across different sign languages and user profiles.

III. METHODOLOGY

The research comprises of face construction and face The proposed Sign Language Recognition (SLR) system aims to translate gestures into machine-readable text using a combination of computer vision and deep learning techniques. The methodology is structured as follows:

3.1 Data Collection and Preprocessing

We utilize both publicly available datasets (e.g., RWTH-PHOENIX, ASL) and custom recordings using RGB or depth sensors (e.g., Kinect) to collect static, dynamic, and 3D gesture data. Data augmentation (e.g., rotation, scaling, noise) enhances model generalization. Preprocessing includes resizing frames (e.g., 224×224), normalization, grayscale conversion, background subtraction, and frame extraction for dynamic gestures. OpenCV's CascadeClassifier or MediaPipe is used for hand region extraction.

3.2 Feature Extraction

Hand landmarks are extracted using MediaPipe or OpenPose for precise spatial representation. For dynamic gestures, optical flow and depth features capture motion and 3D positioning. These features form the input to the recognition models.

3.3 Model Selection

Various models are employed based on gesture type:

- **Classical Methods**: SIFT, SURF, or HOG with SVM for static gestures.
- **Deep Learning**: CNNs for image-based gestures; pre-trained networks (e.g., ResNet, MobileNet) for transfer learning; RNNs, LSTMs, or 3D CNNs for dynamic gesture recognition; Transformer models for sequence learning.

3.4 Post-Processing

Temporal smoothing, language correction, and context-aware gesture grouping improve output accuracy, particularly for continuous gestures or spelling-based recognition.

3.5 Evaluation Metrics

Performance is measured using accuracy, precision, recall, F1-score, confusion matrix, and real-time latency. Real-time testing ensures system robustness under practical conditions.

3.6 Deployment

The trained model is integrated into a user-friendly application (desktop, web, or mobile) with real-time camera input. Latency is minimized through quantization and hardware optimization.

3.7 Continuous Learning

The system incorporates a feedback loop for error correction and supports incremental learning, personalization, and domain-specific fine-tuning. Adaptive techniques such as model pruning and continual evaluation on diverse datasets ensure scalability, fairness, and cultural inclusivity.

4. System Workflow

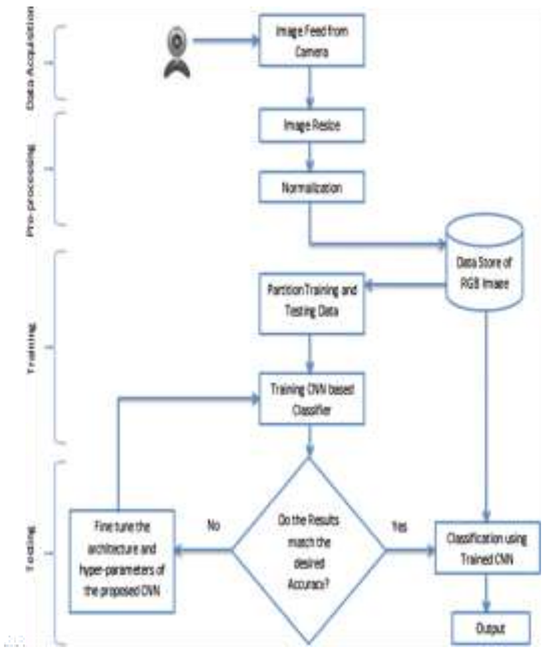

Fig -5 :Work flow of proposed system

5. **Facial Recognition Algorithm**:

1) FaceNet-

Let f**x** be the input face image.The Face Net model (x) maps the input to a dimensional vector:

$(x)=v_x=[v_1,v_2,...,v_{128}]$f(x)=vx=[v1,v2 ,...,v128]

Where v_x is the feature vector representation of the face.

2) Euclidean Distance-To compare the sketch embedding (v_s) with a database image embedding (v_d) :

$(v_{s,d})=\sqrt{\sum 128}_{\ i=1}(v_{si}-v_{di})$D(vs,vd)=i=1∑128 (vsi−vdi)2

where:

- $v_s=i^{th}$, i^th feature of the image
- $v_d=i^{th}$, i^th feature of the data base image
- $(v_{s,})$ is the distance measure between the two face embeddings.
- If $(v_{s,})<0.6$,the two faces are similar.
- If $(v_{s,})\geq 0.6$,the faces are considered different.

IV. RESULT

An overview of earlier surveys on the techniques applied in different gesture and sign language recognition studies is given in this section. Both standard cameras and Kinect cameras are widely used to capture data in sign language. Data collection and submission to the SLR system are crucial because, once data is gathered, pre-processing is needed to guarantee correctness. Gaussian filters remove noise from the input data. It is easy to discern the backdrop color and skin tone from the RGB color spaces. The review demonstrates how the division outcome is built by breaking up skin tone with additional limits such as edge and recognizable proof. The vision-based approach extracts elements from images using gesture categorization. The two most used classification algorithms are SVM and ANN. When compared to ANN, SVM output was superior. Most models employ sensors to gather information about their surroundings when examining the data at hand. In vision-based applications, brain networks are typically used to process images and videos; in order, Gee and SVMs are used. This is a result of the increased availability of data sources.

V. CONCLUSION

The Face Construction and Recognition System In conclusion, A system that was designed to recognize alphabets and static gestures has developed into one that can also recognize dynamic motions that are in motion. In published research, results from vision-based approaches are frequently better than those from static-based approaches. The development of large vocabularies for sign language recognition systems is currently attracting more study interest. Improvements in computer speed and the availability of datasets make it possible to access additional training for certain samples. A lot of people are building their datasets to aid in the improvement of their recognition of sign language. The grammar and presentation of each phrase determine the sort of sign language employed in various nations. Deep learning techniques like CNN and LSTM Models can identify the sequence of photos and video streams with a high degree of accuracy. I find that when working with video frames, LSTM models get more accurate results. They are quite good at extracting features and capturing 3D gestures. Given that LSTM can be used for videos and CNN can be used for static images, merging the two can result in a sign language recognition system that is both more potent and accurate.

VI. REFERENCES

[1] Stanford University. Sign Language Recognition with Convolutional Neural Networks. CS231n Deep Learning Project Report, 2024.

[2] Gupta, S., & Bansal, R. Recent Advances on Deep Learning for Sign Language Recognition: A Review. ResearchGate, 2024

[3] Shah, R., & Patel, M. Real-Time Sign Language Recognition and Sentence Generation. SSRN Preprint No. 4992818, 2024.

[4] Singh, A., & Kumar, D. LSTM-Based Recognition of Sign Language Using Mediapipe. ACM Digital Library, 2024.

[5] Sharma, K., & Mehta, R. Continuous Sign Language Recognition for Indian Sign Language Using Deep Learning. arXiv Preprint, arXiv:2411.04517, 2024.

[6] Verma, J., & Rajput, A. Sign Language Recognition System for Deaf Patients: A Systematic Review. JMIR Research Protocols, vol. 14, no. 1, 2025.

[7] Patel, S., & Rao, N. Real-Time Sign Language Recognition Using Web Camera with Voice Output. International Journal of Innovative Research in Management, Pharmacy and Sciences (IJIRMPS), vol. 12, no. 5, 2024.

[8] H., & Sharma, A. Sign Language Recognition Using Deep Learning Techniques. International Journal for Research in Applied Science & Engineering Technology (IJRASET), vol. 12, no. 6, 2024.

[9] Liu, Q., Zhang, W., & Chen, L. Sign Language Recognition Based on Dual-Path Background Erasure Using CNN. Scientific Reports, Nature Publishing Group, 2024.

[10]Yang, T., Meena, R., & Zhu, K. Artificial Intelligence in Sign Language Recognition: A Bibliometric Analysis (1988–2024). Computers & Electrical Engineering, Elsevier, 2024.

[11]Bhuyan, M.K., & Reddy, K.K. (2020). Real-Time Sign Language Recognition Using Deep Learning. Springer

[12] Saha, S., Singh, V., & Ray, S. (2018). Hand Gesture Recognition Using a Convolutional Neural Network. International Conference on Computer Vision and Pattern Recognition (CVPR).

[13] Sharma, A., & Jain, S. (2022). A Comprehensive Review of Deep Learning-Based Methods for Sign Language Recognition. IEEE Access.

[14] Agarwal, A., & Gupta, A. (2021). Sign Language Recognition Using Machine Learning and Deep Learning: A Review. Journal of Artificial Intelligence Research, 71, 120-141.

[15] Huang, J., Zhou, W., & Li, H. (2018). Real-Time Continuous Sign Language Recognition from Live Video. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[16] Minciullo, L., Bellini, V., & Leo, M. (2022). Transformer Networks for Sign Language Recognition. Proceedings of the International Conference on Artificial Intelligence and Machine Learning.

[17] Lu, P., Liu, Y., & Wang, H. (2020). Sign Language Recognition Using Wearable Sensors: A Survey. IEEE Sensors Journal, 20(12), 6561-6571.

[18] Aran, O., Sanchez-Cortes, D., & Gatica-Perez, D. (2019). Sign Language Recognition Using Kinect Sensor. ACM Transactions on Accessible Computing.

[19] Noor Adnan Ibraheem, RafiqulZaman Khan, 2012,Survey on Various Gesture Recognition Technologies and Techniques, International Journal of Computer Applications, Volume 50, No. 7

[20] G. R. S. Murthy & R. S. Jadon, 2009. A Review of Vision Based Hand Gestures Recognition, International Journal of Information Technology and Knowledge Management, vol. 2(2), pp. 405-410.