Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 15 May 2026
Revised 16 June 2026
Accepted 02 July 2026

Natural Resources for Human Health 2026, Vol. 6 (6s): 283–288

KEYWORDS:
Lorem ipsum, dolor sit amet, consectetur, adipiscing elit, natural resources, phytochemistry.

A Multi-Scale Gabor Filter Bank Approach for Handshape Feature Extraction in Vision-based Sign Language Recognition Systems

Kiran C. Kulkarni[1], Manoj A. Wakchaure[2]

[1]Research Scholar Dept. of Computer Engineering, MET's Institute of Engineering, Nashik Savitribai Phule Pune University, Pune.
kiranc.kulkarni@gmail.com
[2]Research Guide Dept. of Computer Engineering, MET's Institute of Engineering, Nashik Associate Professor, Amrutvahini College of Engineering, Sangamner, Savitribai Phule Pune University, Pune.
manoj.wakchaure@avcoe.org

ABSTRACT: Sign language is the primary mode of communication for millions of Deaf and hard-of-hearing individuals, yet the absence of affordable, accurate, and real-time translation tools continues to isolate this community from mainstream communication channels. This paper reviews the principal issues and challenges confronting vision-based Sign Language Recognition (SLR) systems - spanning data scarcity, linguistic complexity, occlusion, and real-time deployment constraints - and proposes a multi-scale Gabor filter bank for handshape feature extraction as a lightweight alternative to purely deep-learning pipelines. Filter wavelengths (λ) are systematically mapped to kernel sizes using the standard Gabor relation $\sigma = 0.56\lambda$, producing three operating bands - fine (15×15 to 27×27), transition (31×31 to 37×37), and coarse (41×41 to 69×69) - each tuned to a distinct anatomical feature scale, from fingernail edges to full hand and body silhouette. We discuss the design rationale, computational trade-offs, and integration of this multi-scale representation within a broader recognition pipeline, and outline directions for experimental validation.

1. INTRODUCTION

Sign language is a complete, natural language with its own grammar, syntax, and lexicon, used by an estimated 70 million Deaf people worldwide. Despite its linguistic richness, the vast majority of the hearing population has no working knowledge of sign language, and professional interpreters remain scarce, expensive, and rarely available on demand. This asymmetry creates persistent barriers to communication in healthcare, education, employment, legal proceedings, and everyday social interaction.

Automatic Sign Language Recognition (SLR) - the computational task of translating signed gestures into text or speech, and ideally the reverse - has therefore emerged as an important application area at the intersection of computer vision, natural language processing, and assistive technology. Advances in convolutional neural networks, pose-estimation frameworks, and sequence models such as LSTMs and Transformers have made real-time, camera-only SLR increasingly feasible without specialized hardware such as gloves or depth sensors.

Nonetheless, robust SLR remains an open research problem. This paper makes two contributions. First, it consolidates the key issues and challenges reported across the SLR literature into a structured taxonomy spanning data, linguistics, vision, modeling, deployment, and evaluation. Second, it proposes and details a multi-scale Gabor filter bank for handshape feature extraction, designed to capture discriminative hand texture and structure across a spectrum of spatial scales, from millimeter-level fingernail and knuckle detail to full hand and body silhouette.

2. MOTIVATION

The motivation for this work is grounded in the following considerations:

- Communication barrier: The gap between Deaf/Hard-of-Hearing and hearing communities persists across hospitals, schools, workplaces, and public services due to low sign language literacy among the hearing population.
- Interpreter shortage: Certified interpreters are limited in number and cost, and are rarely available in real time or in emergency situations.
- Demand for inclusive technology: Digital accessibility expectations increasingly require that video calls, kiosks, and virtual assistants support sign language natively.
- Technical feasibility: Recent progress in computer vision and sequence modeling has made accurate, low-latency, camera-only SLR realistic for the first time, motivating renewed investment in lightweight and efficient recognition pipelines.
- Two-way communication gap: Most existing systems translate sign-to-text; comparatively little work addresses natural text/speech-to-sign generation.
- Equity and independence: Automated SLR reduces dependency on human interpreters and supports independent access to healthcare, education, and employment, aligning with accessibility mandates such as the ADA and UN CRPD.

3. ISSUES AND CHALLENGES IN SIGN LANGUAGE RECOGNITION

3.1 Data-Related Challenges

- Scarcity of large, standardized, and richly annotated datasets, especially beyond ASL.
- Signer variability in hand size, signing speed, and personal style, which hampers generalization.
- Uneven resource availability across national sign languages (e.g., ISL, BSL, JSL).
- Dataset bias toward controlled lab conditions rather than real-world backgrounds and lighting.

3.2 Linguistic Complexity

- Non-manual markers (facial expression, eyebrow and mouth movement, head tilt) that carry grammatical meaning but are frequently ignored by hand-only systems.
- Co-articulation between consecutive signs, complicating temporal segmentation.
- Distinct grammar and syntax from spoken/written language, complicating gloss-to-text translation.
- Regional and dialectal variation analogous to accents in spoken language.

3.3 Technical and Vision Challenges

- Self-occlusion of fingers, hands, and face during signing.
- Variable background clutter, shadows, and lighting in real-world deployment.
- Motion blur from fast hand movement at standard frame rates.
- 2D depth ambiguity in monocular video, without dedicated depth sensors.
- Latency constraints for continuous, real-time recognition on edge or mobile hardware.

3.4 Modeling Challenges

- Isolated sign recognition is comparatively easy; continuous sentence-level recognition requires robust temporal segmentation.
- Gloss-to-text translation is itself a difficult sequence-to-sequence problem with word-order divergence.

Natural Resources for
Human Health
eISSN: 2583-1194

- Poor scalability from small vocabularies (50-300 signs) to full sign-language lexicons (thousands of signs).
- Effective multimodal fusion of hand shape, motion trajectory, facial expression, and body pose remains open.

3.5 Practical, Deployment, and Evaluation Challenges

- Dependency on gloves, depth cameras, or multi-camera rigs limits real-world deployability.
- Wearable sensors can be intrusive, reducing user acceptance.
- Two-way (text/speech-to-sign) generation via avatars remains less mature than sign-to-text.
- Lack of standardized benchmarks and metrics makes cross-study comparison difficult.
- Limited involvement of Deaf communities in system design and evaluation.

4. PROPOSED METHODOLOGY: MULTI-SCALE GABOR FEATURE EXTRACTION

To address the modeling and vision challenges outlined above - particularly the need to discriminate visually similar handshapes while remaining computationally tractable for near-real-time use - we propose a multi-scale two-dimensional Gabor filter bank applied to hand region-of-interest (ROI) crops. Gabor filters are well suited to this task because their joint localization in space and spatial-frequency captures oriented edge and texture information at a controllable scale, closely matching the multi-resolution structure of the human hand: fine ridge- and edge-level detail near the fingertips, mid-level joint and web-fold structure, and coarse silhouette-level shape.

4.1 Scale-to-Kernel Mapping

For each Gabor wavelength λ (in pixels), the Gaussian envelope standard deviation is fixed at the conventional ratio $\sigma = 0.56\lambda$. The effective spatial support of the filter is taken as 6σ, and the kernel dimension K is the next odd integer above 6σ ($K = 6\sigma + 1$, rounded up), which is required so that the kernel has a well-defined center pixel for convolution. Table 1 summarizes the resulting kernel sizes across thirteen wavelength settings (λ = 4 to 20 px), grouped into three operating bands.

Table 1. Multi-scale Gabor kernel design mapped to target hand and body features.

λ (px)	σ = 0.56λ (px)	6σ (px)	K = 6σ+1	Kernel size	Target body-part feature	Scale band
4	2.24	13.44	15	15×15	Fingernail edge	**Fine**
5	2.80	16.80	17	17×17	Knuckle crease	**Fine**
6	3.36	20.16	21	21×21	Fingertip contour	**Fine**
7	3.92	23.52	25	25×25	Finger boundary	**Fine**
8	4.48	26.88	27	27×27	Wrist crease	**Fine**
9	5.04	30.24	31	31×31	Thumb web fold	**Both**
10	5.60	33.60	35	35×35	Finger joint config.	**Both**
11	6.16	36.96	37	37×37	Palm texture region	**Both**
12	6.72	40.32	41	41×41	Finger gap / palm	**Coarse**
14	7.84	47.04	49	49×49	Full finger column	**Coarse**

λ (px)	σ = 0.56λ (px)	6σ (px)	K = 6σ+1	Kernel size	Target body-part feature	Scale band
16	8.96	53.76	55	55×55	Hand silhouette	**Coarse**
18	10.08	60.48	61	61×61	Forearm / wrist taper	**Coarse**
20	11.20	67.20	69	69×69	Face / body outline	**Coarse**

4.2 Operating Bands

- Fine band (λ = 4-8 px, K = 15×15 to 27×27): targets fingernail edges, knuckle creases, fingertip contours, finger boundaries, and wrist creases - the detail needed to disambiguate visually close handshapes.
- Transition band (λ = 9-11 px, K = 31×31 to 37×37): targets thumb web fold, finger-joint configuration, and palm texture, bridging fine detail and overall shape.
- Coarse band (λ = 12-20 px, K = 41×41 to 69×69): targets finger gaps, full finger columns, hand silhouette, forearm/wrist taper, and face/body outline - the global shape and pose context.

4.3 Orientation and Filter Bank Composition

Each of the thirteen scales is paired with a bank of orientations (typically θ ∈ {0°, 22.5°, 45°, ..., 157.5°}, i.e., 8 orientations) to capture edges and ridges regardless of finger angle. The complete filter bank therefore comprises 13 scales × 8 orientations = 104 Gabor kernels; in practice, a reduced subset (e.g., one representative scale per band × 8 orientations = 24 kernels) can be used for real-time inference, trading discriminative power for latency.

4.4 Computational Considerations

Coarse-band kernels as large as 69×69 are computationally expensive if convolved densely across a full video frame. We therefore restrict Gabor filtering to a hand-bounding-box crop obtained from a lightweight upstream hand detector or pose estimator (e.g., MediaPipe Hands), rather than the full frame. For deployment on mobile or edge devices, kernel separability approximations and FFT-based convolution are recommended to keep per-frame latency within real-time bounds (< 40 ms for 25 fps processing).

5. PROPOSED SYSTEM ARCHITECTURE

The overall recognition pipeline integrates the multi-scale Gabor feature extractor within a broader deep-learning-based sequence model, as follows:

1. Input acquisition: RGB video frames from a standard camera.
2. Hand and pose localization: A lightweight detector (e.g., MediaPipe / OpenPose) extracts hand, face, and body keypoints and bounding boxes per frame.
3. Multi-scale Gabor feature extraction: The proposed filter bank is applied to the hand ROI to produce fine, transition, and coarse texture-and-shape descriptors.
4. Feature fusion: Gabor descriptors are concatenated or fused with CNN-based appearance features and keypoint trajectories.
5. Temporal modeling: A BiLSTM or Transformer encoder models the fused feature sequence to capture co-articulation and continuous signing dynamics.
6. Gloss and text decoding: A connectionist temporal classification (CTC) or sequence-to-sequence decoder maps the encoded sequence to sign glosses and, subsequently, grammatically correct text/speech output.

6. DISCUSSION AND EXPECTED OUTCOMES

The multi-scale Gabor formulation directly targets several challenges identified in Section 3. Fine-scale kernels are expected to improve discrimination between visually similar handshapes that differ only in subtle finger configuration, addressing part of the modeling challenge around vocabulary scalability. Restricting coarse-scale filtering to a detected hand ROI mitigates the real-time latency and computational cost concerns associated with large kernel convolution. However, the approach does not by itself resolve non-manual marker capture, co-articulation segmentation, or gloss-to-text grammatical translation, which remain dependent on the downstream temporal and language modeling stages and on dataset diversity.

7. CONCLUSION AND FUTURE WORK

This paper surveyed the principal issues and challenges in vision-based Sign Language Recognition across data, linguistic, technical, modeling, deployment, and evaluation dimensions, and proposed a multi-scale Gabor filter bank for handshape feature extraction as a computationally efficient complement to deep-learning pipelines. Future work includes empirical validation of the proposed filter bank on standard isolated and continuous SLR benchmarks, ablation studies across scale bands and orientation counts, integration of non-manual marker features, and extension toward two-way (text/speech-to-sign) generation to support fully bidirectional communication.

REFERENCES

[1] I. A. Adeyanju, O. O. Bello, and M. A. Adegboye, "Machine Learning Methods for Sign Language Recognition: A Critical Review and Analysis," Intelligent Systems with Applications, vol. 12, 2021.

[2] A. Wadhawan and P. Kumar, "Sign Language Recognition Systems: A Decade Systematic Literature Review," ResearchGate Preprint, 2019.

[3] T. F. Tao, Y. Zhao, T. Liu, and J. Zhu, "Sign Language Recognition: A Comprehensive Review of Traditional and Deep Learning Approaches, Datasets, and Challenges," IEEE Access, 2022.

[4] J. Huang, W. Zhou, H. Li, and W. Li, "Attention-Based 3D-CNNs for Large-Vocabulary Sign Language Recognition," IEEE Transactions on Circuits and Systems for Video Technology, vol. 29, no. 9, pp. 2822–2832, 2019.

[5] "Advancements in Machine Translation for Sign Language: Approaches, Limitations, and Challenges," Expert Systems with Applications, vol. 213, 2023.

[6] M. Madhiarasan and P. P. Roy, "A Comprehensive Review of Sign Language Recognition: Different Types, Modalities, and Datasets," arXiv preprint arXiv:2204.03328, 2022.

[7] N. Sarhan, N. Abdelhalim, M. El-Nasr, and M. Elhoseiny, "Unraveling a Decade: A Comprehensive Survey on Isolated Sign Language Recognition," in ICCV Workshops (AMFG), pp. 1–19, 2023.

[8] İ. M. Baytaş and İ. Erdoğan, "Signer-Independent Sign Language Recognition with Feature Disentanglement," Turkish Journal of Electrical Engineering & Computer Sciences, vol. 32, no. 3, pp. 420–435, 2024.

[9] D. Fan, M. Yi, W. Kang, Y. Wang, and C. Lv, "Continuous Sign Language Recognition Algorithm Based on Object Detection and Variable-Length Coding Sequence," Scientific Reports, vol. 14, Art. no. 27592, 2024.

[10] A. Desai, M. De Meulder, J. A. Hochgesang, A. Kocab, and A. X. Lu, "Systemic Biases in Sign Language AI Research: A Deaf-Led Call to Reevaluate Research Agendas," in Proc. LREC-COLING 2024 Workshop on the Representation and Processing of Sign Languages, pp. 54–65, 2024.

[11] A. Desai et al., "ASL Citizen: A Community-Sourced Dataset for Advancing Isolated Sign Language Recognition," in NeurIPS 2023 Datasets and Benchmarks Track, 2023.

[12] N. Sarhan et al., "Unraveling a Decade: A Comprehensive Survey on Isolated Sign Language Recognition," in Proc. ICCV 2023 Workshops (AMFG), pp. 1–19, 2023.

[13] D. Fan et al., "Continuous sign language recognition algorithm based on object detection and variable-length coding sequence," Scientific Reports, vol. 14, Art. 27592, 2024.

eISSN: 2583-1194

[14] R. Zuo et al., "Towards Online Continuous Sign Language Recognition via Isolated Sign Dictionary and Sliding Window," in Proc. EMNLP 2024, pp. 11146–11160, 2024.

[15] N. F. Attia et al., "Efficient deep learning models based on tension for sign language recognition," Intelligent Systems with Applications, vol. 16, 2023.

[16] S. Alyami et al., "Reviewing 25 years of continuous sign language recognition: Methods, datasets, and challenges," Information Processing & Management, vol. 61, no. 6, 2024.

[17] S. Srivastava et al., "Continuous Sign Language Recognition System using Deep Learning with MediaPipe Holistic," arXiv:2411.04517, 2024.

[18] S. Das et al., "An Expert System for Indian Sign Language Recognition in Emergency Scenarios," ACM Trans. Asian Low-Resour. Lang. Inf. Process., 2024.

[19] A. Baihan et al., "Sign language recognition using modified deep learning and hybrid optimization," Scientific Reports, vol. 14, 2024.

[20] R. Rastgoo et al., "A survey on recent advances in Sign Language Production," Expert Systems with Applications, vol. 244, 2024.