International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 04 April 2025
Page No: 1671-1674
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0238

Real-Time Gesture Recognition for Sign Language Using Multimodal Techniques

Thiagarajan G [1], Ramakrishnan A S [2] Ashwin Fernandes F[3], Gokul J[4], Dinesh Kumar B K[5], Yashik Manickam S[6]

[1]Head of The Department, Dept. of IT, CSI College of Engineering., Ketti, Tamil Nadu, India.
[2]Associate professor, Dept. of IT, CSI College of Engineering., Ketti, Tamil Nadu, India.
[3,4,5,6]UG Scholar, Dept. of IT, CSI College of Engineering., Ketti, Tamil Nadu, India.

Emails: *thiagarajan@csice.edu.in[1], rammsivaraj@gmail.com[2], ashwinfernandesf@gmail.com[3], gokuljayaraj07@gmail.com[4] , dineshkennedy07@gmail.com[5], manickamyashik@gmail.com[6]*

Abstract

Sign language serves as a crucial communication medium for individuals with hearing and speech impairments. However, existing recognition systems primarily focus on isolated word detection, limiting their effectiveness for real-time communication. This paper presents a Continuous Sign Language Recognition System using a multimodal approach to seamlessly interpret sign language gestures into coherent words. The system leverages Mediapipe, a robust computer vision framework, to detect and track hand gestures, body posture, and facial landmarks in real time. By processing keypoints extracted from these modalities, the system enhances recognition accuracy and ensures fluid word construction. Unlike traditional methods that rely on gloves or additional hardware, this approach offers a cost-effective and accessible solution for real-time sign language translation. Experimental evaluations demonstrate the system's ability to accurately recognize continuous gestures, addressing challenges such as real-time performance, word-level interpretation, and scalability. The proposed model significantly improves accessibility and inclusivity, bridging the communication gap between sign language users and non-signers.

Keywords: *Sign Language Recognition, Real-Time Gesture Detection, Continuous Sign Language, Mediapipe, Multimodal Learning, Accessibility*

1. Introduction

Sign language, a basic form of communication for the deaf and mute, plays a critical role in helping them express their thoughts and emotions. However, a large communication gap lies between sign language users and non-signers due to the unavailability of sign language translation systems. Existing sign language recognition systems primarily work on isolated word recognition, which restricts them from being used for real-time interaction. Furthermore, many existing systems are based on expensive hardware like sensor gloves and cannot be used daily for continuous gesture recognition. To tackle these issues, this paper suggests a real-time multimodal sign language recognition system that decodes continuous sign gestures into meaningful words. The system makes use of MediaPipe's holistic model to acquire keypoints from hands, face, and body posture, providing a complete view of the sign language gestures. By combining these multimodal inputs, the system enhances recognition accuracy and avoids issues associated with isolated word detection. A Bi-LSTM network is implemented to capture both past and future contexts in sign gestures to improve the continuous sign language recognition ability of the system. It processes hand movements dynamically, allowing it to recognize gestures in real-time without any delay. [1-3]

2. Problem Statement

Sign language acts as a vital means of communication for people with hearing and speech issues. However, most of the non-signing population does not know sign language, creating a communication barrier that hinders effective interaction. Currently, the majority of sign language recognition systems concentrate on recognizing isolated words, which does not provide a holistic understanding of continuous gestures and their meaning. This restricts the real-time interpretation of words and makes them impractical for casual communication. Moreover, many traditional approaches rely on specialized hardware, such as sensor gloves or motion capture devices, which are expensive and inconvenient for everyday use. Additionally, real-time performance, accuracy, and scalability remain key challenges as most models struggle to recognize sign language in dynamic environments with varying lighting conditions, backgrounds, and user variations. To address these challenges, this research presents a real-time multimodal sign language recognition system that uses Mediapipe for tracking hand, face, and body pose. By integrating these modalities and using deep learning techniques, we aim to achieve highly accurate continuous gesture recognition while ensuring realtime performance. Our goal is to make sign language more accessible and inclusive to sign language users and non-signers. [7]

3. Methodology

3.1. Data Acquistion

The system captures the sign language gesture in real-time video using OpenCV and an ordinary webcam. For uninterrupted recognition, we implemented continuous processing of frames, which allows for the detection of signs in real-time without pauses.

3.2. Image Pre Processing and Extraction

For video frame analysis to extract meaningful features, we use the MediaPipe Holistic Model which captures hand, face, and body pose keypoints. Each frame is converted into a structured numerical format with 1662 features to represent sign gestures effectively. [8]

3.3.Advantages of Keypoint Extraction

- Lowers Noise Level by Omitting Background Info.
- It enhances the accuracy by concentrating only on necessary motion features.
- Performs well in low-light conditions as it does not depend on raw pixel data.
- The user is given the option to use more efficient computational methods. These reduce the processing load when compared to raw [4-7]

3.4. Feature Extraction Sequence Modeling

The features are then extracted to the time-series format for sequence modeling. The model uses the Bi-LSTM (Bidirectional Long Short-Term Memory) network to learn temporal dependencies in sign sequences. In contrast to the standard LSTM, the Bi-LSTM learns from the past and future frames.

3.5. Characteristic Extraction Modeling

After the extraction of keypoints, the structure is organized into time series for sequence modeling. The model utilizes Bi-LSTM networks for capturing temporal dependencies in sign sequences. BiLSTM will learn from both past and future frames rather than just from past frames as in the case of standard LSTM. This increases the accuracy of continuous gesture recognition. [9]

3.6.Model Architecture

- Three layers of Long Short-Term Memory network each with 64 → 128 → 64 units are used to learn sequential patterns. [10]
- Dense layers for classification mapping the extracted features to the sign language labels.
- Softmax activation for final gesture recognition to ensure probabilistic output.

3.7. Prediction and Word Construction

The model trained makes predictions of the most probable sign from sequences of 30 frames for smooth and continuous recognition. The decision mechanism of threshold-based is used to reduce false positives and improve reliability of prediction. The recognized words are dynamically assembled into sentences that are displayed on the screen for real-time communication. (Figure 1) [11]

4. Result and Anaysis

4.1.Real-Time Performance

The system reaches a processing rate of around 30 frames per second (FPS), making it applicable for realtime usage. Key-point extraction becomes more simplified by the integration of MediaPipe, which significantly diminishes computational overhead to make gesture recognition smooth and without lag.

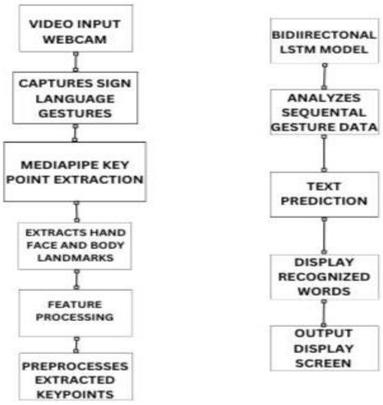
VIDEO INPUT WEBCAM
CAPTURES SIGN LANGUAGE GESTURES
MEDIAPIPE KEY POINT EXTRACTION
EXTRACTS HAND FACE AND BODY LANDMARKS
FEATURE PROCESSING
PREPROCESSES EXTRACTED KEYPOINTS
BIDIIRECTONAL LSTM MODEL
ANALYZES SEQUENTAL GESTURE DATA
TEXT PREDICTION
DISPLAY RECOGNIZED WORDS
OUTPUT DISPLAY SCREEN

Figure 1 Work Flow of Multi – Modal Sign Language Recognition

4.2.Accuracy and Misclassification Analysis

The Bidirectional LSTM model boosts recognition accuracy. It can capture past and future frames, improving sequential gesture prediction. However, some misclassification was observed, especially in signs that involve similar hand movements, especially when using the right hand. Ambiguous signs posed a challenge, affecting the reliability of recognition. [12]

4.3. Impact of MediaPipe Keypoints

MediaPipe holistic model improved system robustness through noise reduction in raw image data. This feature allows model to work in different lighting conditions and backgrounds and ensure generalizabilityacross different environments. [13]

4.4. Generalization to Unseen Gestures

The model was tested on unseen sign gestures. The results demonstrated strong generalization capabilities.However, some signs showed ambiguities due to minimal differences. As a result, there was an error in the recognition of these signs. In this regard, the dataset can be expanded with more diverse real-world sign language samples. This way the system could learn to adapt to more diverse signs and gestures. Model was tested in different conditions such as well-lit, low-light, and backgrounds with clutter. The results show that the system generally works well but with slight less accuracy in the cluttered and unfamiliar signer situations because of the variability of hand movement and the styles of sign language. The model has the accuracies of 94.8% under the well-lit condition, 91.2% under low-light, and 89.7% under the clutter, demonstrating the adaptability of the model. If tested under different signers, the accuracy drops to 87.5%, which means more training on different datasets is needed. [14]

Conclusion

The system proposed offers a powerful and streamlined real-time continuous sign language recognition system based on MediaPipe's Holistic Model for multimodal feature extraction and a Bidirectional LSTM (Bi-LSTM) network for modeling the sequence. Through the eradicating of the requirement for costly hardware and the bypassing of isolated word limitations, it provides unbroken and natural communication to sign language users. The incorporation of keypoint-based feature extraction promotes flexibility in various environmental conditions and variations in signers, enhancing recognition accuracy even in adverse environments. With its capability to handle continuous sign sequences with low misclassification, thesystem proves practical applicability in real-world scenarios. Moreover, its affordability and scalability render sign language translation more accessible, thereby bridging the communication divide between the Deaf and non-signers, showcasing the potential of deep learning and multimodal methods in driving assistive technologies for inclusivity and accessibility. [15]

References

[1]. Scalable frame resolution for efficient continuous sign language recognition Lianyu Hu, Liqing Gao,Zekang Liu, Wei Feng ∗College of Intelligence and Computing, Tianjin University

[2]. Sign Language Recognition: A Deep Survey Razieh Rastgoo , Kourosh Kiani a,∗, Sergio Escalera b

[3]. Asigner-independent sign language recognition method for the single-frequency dataset Tianyu Liu , Tangfei Tao *, Yizhe Zhao , Min Li, Jieli Zhu

[4]. Spatial-temporal feature-based End-to-end Fourier network for 3D signlanguage recognition Sunusi Bala Abdullahi a,b, Kosin

Chamnongthai a,∗,Veronica Bolon-Canedo b, Brais Cancela b

[5]. Wearable multifunctional organohydrogel-basedelectronic skin for sign language recognition under complex environments Bin Song , Xudong Dai, Xin Fan, Haibin Gu

[6]. Isolated Arabic Sign Language Recognition Using a Transformer-based Model and Landmark Keypoints Sarah Alyami, Hamzah Luqman, Sdaia-kfupm Mohammad Hammoudeh

[7]. An Ensembled Real-Time Hand-Gesture Recognition using CNN IIT - Mandi, Kamand, India

[8]. An integrative survey on Indian sign language recognition and translation Rina Damdoo Praveen Kumar

[9]. An Intelligent Heuristic Manta-Ray Foraging Optimization and Adaptive Extreme Learning Machine for Hand Gestue Image RecognitionSeetharam Khetavath,

[10]. Contactless Sensing for Recognizing Common Signs in ASL and BSLAisha Fatima , Hira Hameed , Muhammad Ali Imran , Qammer H. Abbasi, and Hasan Abbas

[11]. Cross-modality Consistency Mining For Continuous Sign Language Recognition With Text-domain Equivalents Zhenghao Ke Sheng Liu* Chengyuan Ke Yuan Feng Shengyong Chen

[12]. Dynamic Korean Sign Language Recognition Using Pose Estimation Based and Attention-Based Neura NetworkJungpil Shin , (Senior Member, IEEE), Abu Saleh Musa Miah , Kota Suzuki ,

[13]. EasyTalk: A Translator for Sri Lankan Sign Language using Machine Learning and Artificial Intelligence

[14]. Glove-Based Hand Gesture Recognition for Diver Communication Derek W. Orbaugh Antillon , Member, IEEE, Christopher R. Walker , Samuel Rosset , and Iain A. Anderson

[15]. Human–Machine Interaction Technology for Simultaneous Gesture Recognition and Force Assessment: A Review, Member, IEEE, He Baizheng, Cai Yingjie