REAL-TIME SIGN LANGUAGE TRANSLATOR

***Project Reference No.:** 49S_BE_2971*

College : Vemana Institute of Technology, Koramangala, Bengaluru
Branch : Department of Information Science and Engineering
Guide(s) : Prof. Jyothi M
Student(s) : Mr. A Nithin Meda
Mr. Abdul Wajeed
Mr. Abhishek G
Mr. Parthasarathi L

Keywords:
Sign Language Recognition, Real-Time Translation, Hand Landmark Detection, MediaPipe, Machine Learning, Random Forest, Computer Vision, Text-to-Speech, Flask, Gesture Classification, Static Gestures, Dynamic Gestures, Accessibility, Assistive Technology.

Introduction:

1. Communication is a fundamental aspect of human interaction, yet individuals with hearing and speech impairments often face significant challenges in expressing themselves to the wider population. Sign language serves as an effective and expressive medium for communication within the deaf and mute community; however, the lack of widespread understanding of sign language creates a communication gap in everyday scenarios. This gap becomes more evident in areas such as education, healthcare, workplaces, and public services, where clear and timely communication is essential.
2. With the advancement of computer vision and machine learning technologies, there is growing potential to develop systems that can automatically interpret human gestures. Vision-based sign language recognition systems, which rely on camera input rather than specialized hardware, offer a practical and cost-effective solution. These systems analyze hand movements and positions to identify sign language gestures and convert them into a form that can be easily understood by non-sign language users.
3. The Real-Time Sign Language Translator project aims to address this communication gap by providing an automated system that translates sign language gestures into readable text and audible speech in real time. The system captures live video input through a webcam, detects hand landmarks, extracts meaningful features, and classifies gestures using trained machine learning models. It supports both static and dynamic gestures and allows users to interact with the translated output through an intuitive web-based interface. By integrating real-time gesture recognition, text generation, and text-to-speech functionality, the proposed system enhances accessibility and promotes inclusive communication.

Objectives:

1. Design and implement an intelligent, vision-based system capable of translating sign language gestures into readable text and audible speech in real time.

2. Reduce communication barriers between individuals with hearing and speech impairments and the general population by providing an automated translation mechanism.
3. Accurately detect and track hand gestures using computer vision techniques and extract meaningful features for gesture recognition.
4. Classify both static and dynamic sign language gestures using machine learning models with minimal latency.
5. Develop a user-friendly web-based interface that supports live video streaming, gesture mode selection, and real-time prediction display.
6. Provide text accumulation functionality that allows recognized gestures to be appended to form words or sentences, and integrate text-to-speech output for improved accessibility.
7. Design a modular and scalable system architecture that supports model retraining and future enhancements such as continuous sign language recognition and multilingual support.

Methodology:

1. The system follows a client-server architecture with four logical layers. The Input Layer captures real-time video using a webcam. The Processing Layer handles hand landmark detection via MediaPipe Hands, feature extraction from landmark coordinates, and gesture classification using a Random Forest classifier trained with Scikit-learn. The Application Layer, implemented using Flask, manages model inference, prediction synchronization, and API routing. The Presentation Layer provides a web-based interface built with HTML, CSS, and JavaScript, enabling live video display, mode selection, prediction output, and text-to-speech functionality.
2. Each video frame undergoes preprocessing before hand landmarks are extracted. The 21 hand joint positions (x, y coordinates) detected by MediaPipe are converted into numerical feature vectors. For static gestures, a Random Forest model trained on spatial landmark features classifies individual alphabet signs. For dynamic gestures, temporal sequences of landmark frames are analyzed to recognize multi-frame sign expressions such as 'I Love You' and 'Thank You'. Predictions are streamed to the frontend via MJPEG, and translated text can be accumulated and converted to audible speech using the browser's Web Speech API.

Result and Conclusion:

1. Experimental testing of the Real-Time Sign Language Translator demonstrated reliable recognition of both static and dynamic sign language gestures with low latency on standard consumer hardware. Static alphabet gestures including letters such as 'O', 'C', and 'B' were recognized with high accuracy using spatial hand landmark features. Dynamic gestures including 'I Love You' and 'Thank You' were successfully detected by analyzing temporal hand motion patterns. The system operated smoothly without requiring high-end GPUs, validating the effectiveness of the lightweight Random Forest classifier approach over computationally expensive deep learning models.
2. In conclusion, the Real-Time Sign Language Translator successfully demonstrates a practical, accessible, and deployable solution for bridging the communication gap between hearing-impaired individuals and the general population. The integration of real-time gesture recognition, text output, and text-to-speech functionality in a web-based interface makes the system suitable for real-world assistive communication

applications. Its modular design supports future enhancements including continuous sign language recognition, multilingual support, facial expression analysis, and mobile deployment.

References:

[1] V. Bazarevsky et al., "MediaPipe Hands: On-Device Real-Time Hand Tracking," Google Research, 2020.

[2] R. K. Pathan et al., "Sign Language Recognition Using the Fusion of Image and Hand Landmarks through Multi-Headed CNN," IEEE Access, 2023.

[3] P. Poladiya et al., "Sign Language Detection Using Deep Learning," International Journal of Engineering Research, 2024.

[4] A. S. Selvi et al., "Real-Time Sign Language Detection Using Deep Learning," IEEE Access, 2025.

[5] N. Aloysius et al., "Optimized Multi-Modal Conformer-Based Framework for Continuous Sign Language Recognition," Pattern Recognition Letters, 2025.

[6] T. Tao et al., "Sign Language Recognition: A Comprehensive Review of Traditional and Deep Learning Approaches," ACM Computing Surveys, 2024.

[7] S. Rautaray and A. Agrawal, "Vision-Based Hand Gesture Recognition for Human-Computer Interaction," Artificial Intelligence Review, 2015.

[8] D. Molchanov et al., "Online Detection and Classification of Dynamic Hand Gestures," CVPR, 2015.

Future Scope:

The future scope of this project includes several directions for enhancement. The system can be extended to support continuous sign language recognition for sentence-level interpretation instead of isolated gestures. Incorporating facial expression and body posture analysis can improve contextual understanding. The system can be expanded to support multiple sign languages for use across different regions and communities. Integration with mobile platforms through a dedicated application would enhance accessibility and portability. Advanced language models can be incorporated for grammar correction and sentence formation. Additional improvements such as cloud deployment, adaptive learning based on user feedback, and robustness against varying lighting conditions and backgrounds would further enhance scalability and real-world deployment in educational institutions, public service environments, and assistive communication platforms.