Deep Science Publishing
https://doi.org/10.70593/978-81-983916-5-0_4

Chapter 4

Hand gesture recognition for sign language translation

Jithu Varghese Jacob [1], Anwin K Biju [2], Rajesh Kanna R [3]

[1,2,3] *Department of Computer Science, Christ University, Bengaluru, India.*

Abstract: This project aims to develop software that translates Indian Sign Language (ISL) hand gestures into real-time text and speech. A custom symbol module enables users to add new gestures, enhancing user experience. Voice call integration allows real-time gesture-to-speech translation during calls. Optimized machine learning models ensure efficiency and accessibility, creating a solution for inclusive communication. The application relies on optimized machine learning models from TensorFlow Lite which enables safe gesture recognition across different mobile devices. The application needs this specific technology due to its power to maximize accessibility and usability because it operates well on devices with limited processing capacity. The system's architecture puts inclusivity first with its purpose to create an accessible communication tool which serves the deaf and hard-of-hearing community through user-friendly functions. The implementation of thorough error handling and feedback functions in the software system creates improved user experience along with better system dependability. The initiative resolves the current weaknesses of sign language translation solutions through its personalized and efficient software which is accessible to users. A modernized translation framework creates new possibilities for better quality of life among ISL users because it promotes fluid and natural speaking opportunities across different situations. This system includes scalability features to enable future adjustments according to developing requirements.

Keywords: Customizable gestures, gesture recognition, Indian Sign Language, machine learning, real-time communication, voice call integration.

1.1 Introduction

The main mode of communication among the Deaf and Hard of Hearing is sign language. The translation of sign language into text or speech remains quite limited in many applications, such as real-time voice calls. The proposed project will create a mobile application to translate Indian Sign Language gestures into text and speech in real time, thereby bridging the gap between sign language users and non-sign language users. This project allows users to personalize the application using new symbols for words that are not included in the standard dataset. Recording short videos of new gestures makes the application automatically ex- tract key frames; this makes the multiple image capture process automatic, thus making it smooth and friendlier to use. Optimized mobile device models applied to machine learning will have a flexible framework that may enable the predefined user vocabulary.

Other than gesture translation, the application will offer voice calls and con- vert sign language into synthesized speech in future update. This will make it easier for a person with sign language to communicate real-time in two ways with someone with spoken language. This allows interactions to be more accessible and inclusive. They are using TensorFlow Lite for gesture recognition, signalizing the state of communication and WebRTC to make voice calls feel natural in sign language conversations. The technology stack, system architecture, and implementation strategy for an innovative tool that would completely eliminate the existing limitations of trans- lation of sign language to speech, improve user experience, and reduce computational costs.

1.2 Literature review

(Ebey Abraham et al, 2019) It designed the system in order to bridge the gaps of communication between deaf and others through making ISL into speech translation. For this system, the hand gestures were detected through a sensor glove containing flex sensors and IMU. Classified data have been collected which were classified through LSTM networks, which are efficient learners of long-term dependencies. It may classify 26 gestures with an ac- curacy of 98%. Thus, this proves that LSTM based neural networks have real-time capability for the translation of ISL.

(Harini R et al, 2020) proposed computer vision and machine learning work that translates sign language into text. It captures a sign gesture using a we- bcam. Subsequently, it preprocesses the sign through background subtraction, while classification is done through a CNN. The model does pretty well with high accuracy at 99.91%. No equipment such as gloves and sensors is needed, but it relies on computer vision to understand the gestures.

Mohammed Elmahgiubi et al, 2015) proposed a wearable device in the form of a smart glove, equipped with sensors, that translates sign language gestures to readable text. The glove, equipped with flex sensors, contact sensors, and accelerometer, captures hand gestures to send data for interpretation to the microcontroller. The textual output might be displayed at an LCD screen or forwarded to a smartphone. It recognizes 20 out of 26 letters of American Sign Language with 96% accuracy. It is simple and compact and cost-effective in design.

(Zeyu Liang et al, 2023) summarizes a vast array of current sign language trans- lation approaches along with their techniques. Ch. outlines challenges in- volved together with advances in converting it to spoken or written out- put via machine learning. This work splits SLT into four significant tasks: Sign2Gloss2Text, Sign2Text, Sign2(Gloss+Text), and Gloss2Text. More so, the paper outlines transformer-based architectures applied in SLT and discusses some of the major challenges such brings: scarcity of data and complexity in grammar of sign languages. Lastly, the survey is summed up by proposing directions of future work to make improvements on the SLT model (Sign Language Translation).

Necati Cihan Camgoz et al, (2018) proposed Neural SLT as a new distinct task from SLR. Therefore, the objective of SLT is the direct translation of videos from sign language into spoken language text directly, capturing the differ- ence in the grammar and word orders of both languages. Authors propose an encoder-decoder model using convolutional and recurrent neural networks with attention mechanisms. They contribute a continuous SLT data-set called RWTH PHOENIX-Weather 2014T for the German Sign Language, and state baseline results for several architectures of SLT.

Mary Jane C. Samonte et al, (2022) looks at deep learning models of translation from sign language into text, aimed at increasing accessibility in communication in cases where the concerned person may either be speech and hearing impaired. Here, the importance of CNNs, CTC, and DBN concerning text production when interpreting hand gestures is brought to attention. A deep learning-based proposed model was developed with respect to careful review of all relevant studies for the precise recognition of signs and language processing. Such enhanced systems of translation would minimize communication barriers as well as barriers of accessibility by the deaf and hard-of-hearing community.

Adria´n Nu´ n˜ ez-Marcos et al, (2023) offered a holistic view of methods, challenges, and advances in sign language translation. Traditional approaches include rule-based methods and statistical machine translation. Recently, there also appeared a great advance, using neural models based on deep learning techniques, referred to as NSLT; a good quality of such robust datasets, like the mentioned RWTH-PHOENIX-2014T, is also one more requirement in training its suitable machine learning model using correct results. In this survey, it can be seen how gesture segmentation, gloss translation, and regional sign language variation will be further advanced to sup- port smooth translation with precision across languages.

1.3 Methods and materials

A. Tools and Technologies

• MediaPipe: MediaPipe is a highly optimized library for real-time hand tracking and gesture detection.It gives you the ability to accu- rately track hand and finger move- ments by providing precise hand landmark detection. It would be a great benefit for any ISL (Indian Sign Language) applications, since the scenerio of detecting gestures on both Mobile and Desktop is persistent.

• TensorFlow Lite: Tensor Flow Lite is used for deploy- ment of cus- tom trained model to recognize gestures on mobile. TensorFlow Lite framework provides an op- timized runtime for machine learning mod- els that is extremely useful in keeping performance and reducing la- tency on resource-constrained mobile devices.

• OpenCV: It is used to add new gestures in capturing and processing video frames.5 OpenCV can automatically take frames from a video that saves the user considerable work when adding custom gestures as he need not capture several still images.

• Google TTS API & pyttsx3: This API offers the facility for on-cloud speech synthesis. It converts ISL gestures recognized to live audible speech. Good performance and high-quality output that would be perfect for developing an ISL accessible application is maintained with multilin- gual support, making the output surprisingly very natural.

• WebRTC: We have implemented WebRTC (n.d.)-a real-time communication framework that allows low-latency audio/video streaming-for real-time integration of voice calls. As such, voice calls are seamlessly integrated into the sign language translation application, allowing natural and fluid communication.

• Keras: Keras (n.d.) is a user-friendly, high-level API for TensorFlow that allows for easy development and training of machine learning models. The simplicity of Keras allows fast prototyping and experimentation with different model architectures and training parameters.

The system architecture has been designed keeping modularity and scalability in mind, and the entire system is divided into three big modules:

The Hand Tracking and Gesture Recognition Module utilizes MediaPipe for real-time hand tracking and a custom-trained TensorFlow Lite model for gesture recognition. It obtains video input from the camera of the device, hand landmarks are being detected, and gestures are classified. The output of this module consists of a continuous sequence of recognizable gestures.

Text Generation Module translates the sequence of recognized gestures into text by mapping them to corresponding graphical representations based on a user's custom vocabulary. It contains error handling and correction mechanisms to increase text-generation accuracy.

This module works to guarantee that the translated text is presented to the user in an accessible and user-friendly mode: text is being converted to audible speech via the Google TTS API and pyttsx3.

The implementation strategy will embark on the iterative development paradigm, starting with a Minimum Viable Product (MVP) that will accommodate the critical functionalities. In the subsequent iterations, more advanced functionalities will be introduced, such as improved gesture segmentation, data augmentation, and ensemble methods. User feedback is continuously incorporated into the development process to ensure the fulfillment of ISL users' needs.

The evaluation of the ISL translator's performance will be rigorous, carried out by employing a plethora of quantitative and qualitative metrics.

The quantitative evaluation will engage in testing the accuracy of gesture recognition based on standard metrics, which include precision, recall, F1 score, and accuracy. For this purpose, the speed and performance of the system will be assessed in terms of FPS and latency, giving highly objective measures for the evaluation of performance.

In the qualitative evaluation part, user feedback will be collected from surveys and interviews for evaluating usability, user experience, and satisfaction. Those insights will underline the strengths and weaknesses of this system from the perspective of the target users.

The future development will focus on the key points below:

Expand the Vocabulary: Expand signs under ISL that can be referenced by the system by collecting and annotating a big dataset.

Improve Robustness: The robustness of the gesture recognition model shall be improved through data augmentation and ensemble methods so that variations in lighting, positioning of the hand, and individual signing styles can be catered for.

Integrating Other Features: Accessibility will be further enhanced by being equipped with other features like being sentient to translate between spoken languages, support regional variations of ISL, and integrate with various other assistive technologies.

Advanced Deep Learning Techniques: The application of transformer-based models and other recent innovative deep-learning techniques could be evaluated critically to boost the accuracy and efficiency of translation.

Building a Larger Dataset: A greater component will be dedicated to making a more varied dataset of ISL gestures through enhancement by collaboration with the ISL community to ensure their complete representation.

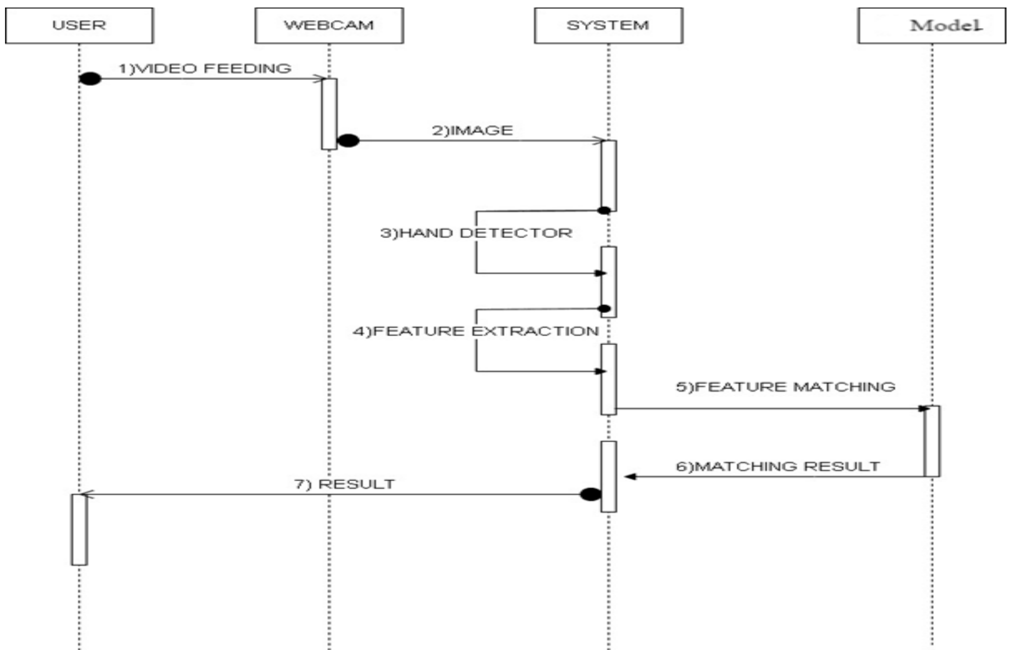

Fig. 1. Sequence Diagram

B. Model Training

- Utilize TensorFlow and Keras for model development.
- Employ transfer learning and data augmentation to im- prove robustness.
- Optimize for mobile environments using TensorFlow Lite.

1.4 Results and discussions

Real-Time ISL to Text and Audio Translation: The project will enable immediate translation of ISL gestures into text and spoken language. This outcome will enhance com- munication for ISL users, bridging the communication gap with non-signers in various environments.

• Custom Gesture Recognition: Add custom gesture for user's vocabulary and particular use. This personal in- terface would help users to configure the system without demanding advanced technical know-how.

• Integrated TTS with Cloud and Offline Options: This will be an application featuring integration of TTS function both with Cloud and offline Options. It will thus use the offline TTS version, which is base upon pyttsx3 while supporting translation functionalities even on cases of connectivity.

• Scalable Gesture Recognition Model: This project will deliver an adaptive model, which, as the vocabulary of ISL continues to grow, will take into consideration new gestures. That flexibility will continuously allow it to be evolved to be strong and scalable.

• Greater Inclusivity for ISL Users: Providing a mobile application that will translate ISL is aimed at opening access to education, public services, social interaction, and voice communication. Through this project, more inclusive conditions for these activities can be attained.

• Broader Impact on Accessibility: These technologies con- tribute significantly to accessibility as they promote the use of assistive technology for the ISL users and have opened up access to different social structures.

Conclusions

The proposed project will introduce the entire approach of developing a real-time Indian Sign Language translation for PC application. The application is designed to offer accessible choices of communication for the users of the ISL. It will successfully translate sign language into text and speech using gesture recognition, machine learning, and TTS technologies. This is the feature that allows the model to add custom gestures, thereby catering to the needs of individual vocabularies without requiring much retraining to remain adaptive and responsive. It further brings out the user-centered design aspect, thus easy to use and efficient in PC-based applications. Through accessible tools such as TensorFlow, MediaPipe, and TTS libraries, this application brings about the possibility of having a cost-effective yet high-performance ISL translation tool. This is why this project is critical in filling the communication gap for ISL users and promoting inclusivity in digital interactions and paving the way for further developments in assistive technology for sign language translation.

References

Abraham, E., & Others. (2019). Real-Time Translation of Indian Sign Language Using LSTM.

Camgoz, N. C., & Others. (2018). Neural Sign Language Translation.

Chung, J., & Lee, D. (2021). Hand Pose Estimation for Robust Gesture Recognition in Complex Environments. *Journal of Machine Learning Research*.

Elmahgiubi, M., & Others. (2015). Sign Language Translator and Gesture Recognition.

Goyal, S., Pandey, P., & Bhatia, N. (2021). Real-Time Sign Language Recognition Using Mediapipe and TensorFlow. *International Journal of Advanced Computer Science and Applications*.

Harini, R., & Others. (2020). Sign Language Translation.

Koller, O., Camgoz, N. C., & Ney, H. (2019). Weakly Supervised Learning of Sign Language Interpretations via Parallel Corpus Mining. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Liang, Z., & Others. (2023). Sign Language Translation: A Survey.

Ronchetti, F., Quiroga, F. M., Estrebou, C., Lanzarini, L., & Rosete, A. (2023). Sign Language Recognition Without Frame-Sequencing Constraints: A Proof of Concept on the Argentinian Sign Language. *arXiv preprint arXiv:2310.17437*.

Samonte, M. J. C., & Others. (2022). Deep Learning in Sign Language Translation.

Shi, Y., Yang, L., & Feng, X. (2023). Attention-Based Encoder-Decoder Models for Sign Language Translation. *Neural Computing and Applications*.

Shukla, K., & Jain, R. (2022). Hybrid Models Combining CNN and LSTM for Sign Language Gesture Recognition. *Springer Advances in Neural Networks and AI*.

Sun, D., & Li, Z. (2022). A Review of Gesture Recognition Techniques for Human-Computer Interaction. *IEEE Access*.

Thomas, C., & Ahmad, T. (2020). Exploring Sign Language Translation Challenges Using Transfer Learning. *Computer Vision and Image Understanding*.

Wang, P., & Chen, F. (2023). Open-Source Solutions for Real-Time Sign Language Translation. *Proceedings of the ACM Symposium on Applied Computing*.