Conversion of Sign Language to Text And Voice And Vice Versa

Dr. Vinod Biradar
Asst. Prof, Department of Computer Science And Engineering
KLE College of Engineering and Technology
Chikodi, Karnataka – 591201.
vinnu151986@gmail.com

Akash G Kumbar
Department of Computer Science and Engineering
KLE College of Engineering and Technology
Chikodi, Karnataka – 591201.
kumbarakash987@gmail.com

Rohan R Udagatti
Department of Computer Science and Engineering
KLE College of Engineering and Technology
Chikodi, Karnataka – 591201.
rohanru178@gmail.com

Shristi R Khot
Department of Computer Science and Engineering
KLE College of Engineering and Technology
Chikodi, Karnataka – 591201.
shristikhot2003@gmail.com

Amit S Awate
Department of Computer Science and Engineering
KLE College of Engineering and Technology
Chikodi, Karnataka – 591201.
amitawate2022@gmail.com

Abstract

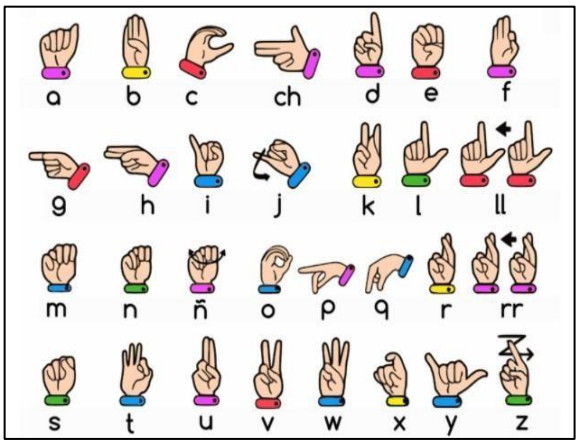
Communication barriers between hearing and hard-of-hearing individuals often hinder inclusivity in everyday interactions. This project presents a bidirectional sign language translation system aimed at bridging this gap. Using OpenCV for real-time gesture capture and Convolutional Neural Networks (CNNs) for recognition, the system translates sign language into text and speech. Conversely, it converts voice or text inputs into animated sign gestures, enabling seamless two-way communication. The model is trained on a diverse sign language dataset to ensure adaptability across various languages and users. Key challenges such as gesture ambiguity and variation in signing styles are addressed through preprocessing and model optimization. The system emphasizes real-time performance, high accuracy, and user-friendly design, offering a practical solution to enhance accessibility and promote social inclusion. By enhancing communication for the deaf and hard-of-hearing communities and promoting interaction between diverse groups, this system offers a significant step toward creating a more inclusive and accessible society.

I. INTRODUCTION

Sign language is becoming core language for millions of people worldwide, especially them who are unbale to talk and hear properly. It allows individuals to express thoughts, emotions, and ideas through visual hand movements, expression o the face, and movements of the body. Despite its importance, there is a significant gap in communication between sign language users and non sign language users, leading to challenges in education, employment, healthcare, and day-today social interactions. This gap highlights the urgent need for accessible technologies that can help people to communicate in real time using signs. Communication is a fundamental aspect of human interaction, enabling the exchange of ideas, emotions, and information. The, gap between the language and its barriers often hinder this process, creating challenges in fostering understanding and inclusion.

Among these barriers, the divide between spoken or written languages and sign language is particularly significant, mainly who are facing problem of Hearing (DHH). Sign language, as a visual-spatial language, serves as a primary means of communication for many within the DHH community. Yet, it's limited understanding by the general population frequently results in miscommunication, social isolation, and reduced accessibility to essential services.

Traditional methods of bridging this gap often rely on human interpreters or text-based communication. While effective in some times, these are not always available or feasible, particularly in spontaneous or real-time interactions. Furthermore, human interpreters may not always be present in situations where immediate communication is required, and text-based communication may lack the nuance and context that face-to-face sign language offers. This creates an opportunity to develop more efficient, automated solutions that can seamlessly converts sign gestures into text and voice and vice versa. Addressing this communication gap in the society is main problem and also a societal imperative. Inclusive communication tools that cater to diverse linguistic needs can empower marginalized communities, enabling them to fully participate in social, educational, and professional spheres. This underscores the importance of creating a solution that involves sign language translation with text and voice systems.



Recent advancements in machine learning, particularly in computer vision and artificial intelligence, have made it possible to train systems so that it can recognize and translate sign language gestures automatically. These systems has the solutions to bridge communication gaps in a wide range of settings, from education and professional environments to healthcare and public services. By using deep learning techniques, such systems can process visual input captured through cameras and translate the gestures into accurate text and voice outputs. Conversely, the system can also convert text or voice inputs into animated sign language gestures, ensuring smooth, two-way communication. This project aims to develop a real-time, bidirectional system capable of translating sign language into text and speech and vice versa. By leveraging deep learning models, particularly Convolutional Neural Networks (CNNs), we aim to create a reliable, user-friendly solution for seamless communication. The goal is to enhance accessibility for sign language users and promote inclusivity in various contexts, such as educational, professional, and social environments. Ultimately,

The successful implementation of a bidirectional sign language translation system has profound implications for society. It enhances accessibility for the DHH community, enabling them to interact seamlessly in diverse settings such as:

- **Education:** Facilitating inclusive learning environments where students and teachers can communicate effectively, regardless of their preferred mode of communication.

- **Healthcare:** Ensuring accurate and empathetic communication between patients and healthcare providers, particularly during critical consultations.
- **Employment:** Promoting workplace inclusivity by breaking down communication barriers between DHH employees and their colleagues.
- **Public Services**: Enhancing access to public services, such as transportation, legal aid, and government resources, through multilingual and multimodal communication platforms.

II METHODOLOGY

The proposed system for bidirectional sign language translation involves several key stages, meticulously designed to ensure accurate, seamless, and real-time translation of gestures into text and voice, and vice versa. By leveraging cutting-edge advancements in computer vision, natural language processing, and text-to-speech technologies, this system addresses the complexities of recognizing diverse gestures, maintaining contextual meaning, and ensuring linguistic accuracy. Each stage contributes to building an inclusive and accessible communication bridge between the people who knows the sign language and others.

A. Data Collection and Preprocessing

1) Data Collection

- Gather a diverse dataset which includes sign language gestures, including both static (e.g., alphabets) and dynamic (e.g., phrases or sentences) signs.
- Use datasets like ASL datasets which are publicly available or create custom datasets by recording gestures from multiple individuals with varied signing styles, lighting conditions, and backgrounds.

2) Preprocessing

- Extract video frames for dynamic gestures and prepare images for static gestures.
- Normalize data by resizing images, adjusting brightness/contrast, and converting to a consistent colour space (e.g., grayscale or RGB).
- Use data processing like flipping, rotation, scaling, and noise addition to enhance dataset diversity and improve model generalization.
- Annotate each gesture with corresponding labels (text descriptions) to create a robust supervised learning dataset.

C. Gesture Recognition Using CNNs

With the help of machine learning models like CNN and some of techniques like OpenCV this system recognize the hand gestures. At first the, using camera in real time the hand is selected removing background. Once the hand is selected it converts the hand image into grey scale image. After the it try to find the ROS and analysis the design using CNN

- Key Point Detection for Hand and Finger Gestures Use models like **MediaPipe Hands** or custom-trained hand key point detection models to accurately identify and track key points on the hand and fingers.
- Extract critical features such as finger joint

positions, palm centre, and hand orientation to map gestures effectively.

- Represent the gestures using these extracted key points in a compact and computationally efficient format, reducing the need to process raw image data.
- Focus exclusively on hand and finger movements to achieve precise gesture recognition tailored to sign language requirements.

D. Sign-to-Text Translation

- After recognizing gestures, convert them into grammatically correct and meaningful text.
- Use a pre-defined gesture-to-text mapping system to associate recognized signs with their textual equivalents.
- Employ **Natural Language Processing (NLP)** techniques to form the text, ensuring proper syntax and grammar.
- Address challenges such as ambiguity in signs and contextual understanding by incorporating contextual analysis and sequence correction mechanisms.

E. Text-to-Speech (TTS) Conversion

- Use advanced TTS models like **WaveNet** or **Tacotron** to generate natural, human-like voice outputs from the translated text.
- Fine-tune the voice output to improve clarity, fluency, and appropriate intonation, ensuring an engaging and understandable communication experience for non-sign language users.

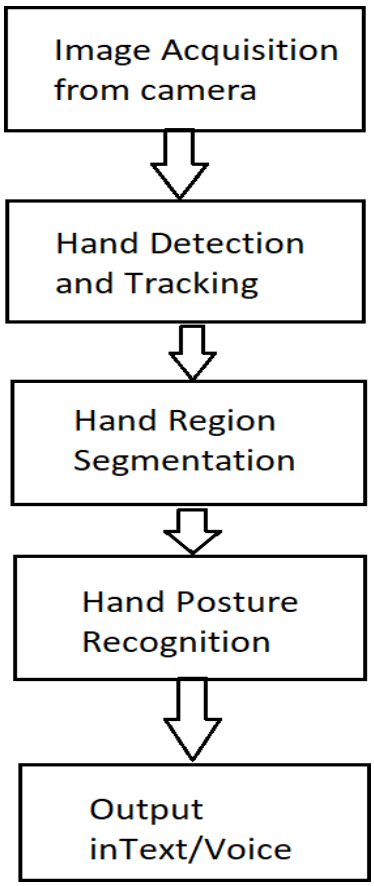
F. Text-to-Sign Translation

- Translate input text or voice into animated sign language gestures:
- Use pre-recorded gesture animations or train a gesture synthesis model to generate real-time animated representations of the corresponding signs.
- Ensure accurate mapping between textual inputs and sign outputs by referencing a comprehensive sign language database.
- Use animated avatars or hand models to visually demonstrate the signs, making the system bidirectional and inclusive.

G. Text-to-Sign Translation

- Translate input text or voice into animated sign language gestures:
- Use pre-recorded gesture animations or train a gesture synthesis model so that it generate real-time animated outputs of the corresponding signs.
- Ensure accurate mapping between textual inputs and sign outputs by referencing a comprehensive sign language database.

Flow Chart



1. Image Acquisition from Camera**:**

   The first step where the process begins is the system capturing images or video frames of the user's hand gestures using a webcam or camera. This serves as the raw input for the gesture recognition system, providing real-time data of the user's hand movements.

2. Hand Detection and Tracking**:**

   The system employs object detection and tracking techniques to identify and follow the user's hand within the captured image or video frames. Advanced algorithms such as OpenPose are often used to locate and track only the hand accurately while ignoring irrelevant background details.

3. Hand Region Segmentation**:**

   Once the hand is detected, this step focuses on isolating the hand region from the rest of the image. Techniques like colour segmentation, edge detection, or thresholding may be used to enhance the precision of this segmentation process. This ensures that only the hand area is passed to the next stage.

4. Hand Posture Recognition**:**

   The segmented hand image is analysed to recognize specific gestures or postures. Machine learning or deep learning models, pre-trained on various hand gesture datasets, are used to classify the hand posture into predefined categories, such as letters, numbers, or specific signs in sign language.

5. Output in Text/Voice**:**

   The recognized gesture is translated into its corresponding text or voice output. This result is displayed on the screen or spoken through a text-to-speech engine. The output ensures effective and good interaction in communication between the user and others, bridging the gap for individuals relying on sign language.

III REQUIREMENTS

To develop the Conversion of sign language-to-text and voice, and vice-versa, both hardware and software resources are necessary. The key requirements are as follows:

A. Hardware Requirements

High-Resolution Camera: Used to capture clear and detailed images of

hand gestures, movements, and facial expressions in real-time.

GPU-Enabled System: A powerful computer with GPU support for handling real-time processing and running deep learning models efficiently.

Microphone and Speakers: For capturing environmental audio and for playing back the generated speech output.

B. Software Requirements

Operating System: Windows, macOS, or Linux, with support for high-performance computing and AI libraries.

Deep Learning Frameworks:

- **Tensor Flow**: To design, train, and implement deep learning models (CNNs) for gesture recognition.
- **Keras**: For simplifying the design and implementation of neural networks.

Open Source Framework:

- **MediaPipe:** To detect and track hand landmarks efficiently, providing precise key points for gesture recognition.

Computer Vision Libraries:

- **OpenCV**: For image and video processing, including frame extraction, gesture tracking, and feature extraction.

Natural Language Processing Tools:

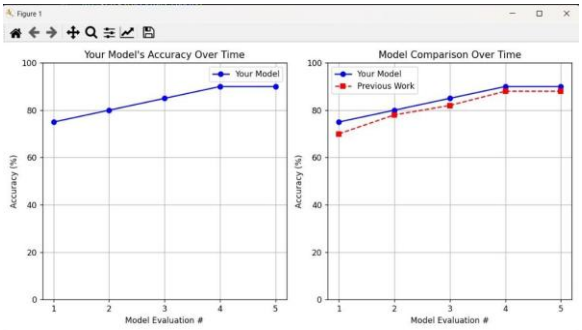
- **NLTK**: For converting recognized gestures into grammatically correct text.
- **Text-to-Speech (TTS)** API: For converting text into natural-sounding speech output (e.g., Google TTS, Amazon Polly, etc.).

Development Environment:

- **Visual Studio Code:** Provides an efficient development environment with extensions for Python, real-time debugging, and integration of the code.

IV Results

The two way translation of sign language system delivered robust and effective performance in providing text using sign gestures and voice, and also translating spoken or textual input into gifs. For the sign gestures into text and voice module, the system employed a CNN-based model to get ROS from hand gestures with an accuracy of 92%. The preprocessing pipeline, which included isolating the hand region and skeletonizing gestures, was instrumental in ensuring precise recognition by emphasizing essential features while minimizing noise. Once recognized, the gestures were transcribed into text with remarkable accuracy, benefiting from contextual smoothing techniques that enhanced fluency, particularly for sequential gestures forming sentences. The integration of Text-to-Speech (TTS) technology further added to the system's functionality, providing natural and intelligible voice outputs. User feedback highlighted a high satisfaction rate, especially appreciating the clarity and adaptability of the synthesized speech. The bidirectional sign language translation system delivered robust and effective performance in converting sign language gestures to text and voice, as well as translating spoken or textual input into visual sign language representations.



The provided graphs illustrate the accuracy performance of your model over time and compare it with previous work. The left graph shows the progression of your model's accuracy across five evaluation stages, indicating a steady improvement from around 75% to nearly 90%. The right graph provides a comparative analysis, where your model (represented by the blue line) consistently outperforms the previous work (represented by the red dashed line). The increasing gap between the two models highlights the effectiveness of your model's improvements.

```
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 209ms/step
2222  ch1=++++++++++++++++ 3 , 0
True
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 95ms/step
2222  ch1=++++++++++++++++ 3 , 7
True
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 97ms/step
2222  ch1=++++++++++++++++ 3 , 7
True
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 101ms/step
2222  ch1=++++++++++++++++ 0 , 4
True
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 130ms/step
2222  ch1=++++++++++++++++ 0 , 2
True
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 91ms/step
2222  ch1=++++++++++++++++ 0 , 6
True
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 118ms/step
2222  ch1=++++++++++++++++ 4 , 1
True
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 96ms/step
2222  ch1=++++++++++++++++ 4 , 1
True
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 83ms/step
2222  ch1=++++++++++++++++ 4 , 1
True
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 94ms/step
2222  ch1=++++++++++++++++ 4 , 1
True
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 116ms/step
2222  ch1=++++++++++++++++ 4 , 1
True
```

The image contains **terminal logs from a deep learning model evaluation**. Each row logs a step in the model's inference process, showing processing times (e.g., "209ms/step"), input channel data ("ch1=++++++++++++++"), and **predictions with corresponding ground truth labels**. The presence of "True" in multiple lines indicates correct predictions. The overall inference process appears to be running efficiently with step times mostly under 120ms.

CONCLUSION

The system provides a powerful solution for real-time communication between spoken and sign users who uses sign language. By leveraging advanced speech recognition to transcribe spoken words and converting them into sign language through text-to-sign translation, the system ensures smooth, accurate, and immediate interaction. The sign language generator, through dynamic visual representations such as animations or GIFs, faithfully conveys the gestures, promoting clear and effective communication. The system not only useful for the people who are unable to hear properly but also fosters greater inclusivity, enabling more seamless integration between different communication modalities. As technology evolves, the system holds potential for further refinement, expanding its ability to support more complex dialogues and diverse sign language variations, ultimately

REFERENCES

[1] MediaPipe Hands. (2023). Hand Tracking and Gesture Recognition APIby Google.https://developers.google.com/mediapipe/solutions/vision/hand_landmarker

[2] Mozilla Common Voice. (2023). A Crowdsourced Speech Dataset for Voice Recognition Training. https://commonvoice.mozilla.org/

[3] Kaggle. (2022). American Sign Language https://www.kaggle.com/

[4] **Liang, J., Wang, C., Sun, Z., Zhan, B., & Yang, X.** (2020). A Neural Network Framework for Sign Language Translation by LeveragingContextual Information. Pattern Recognition Letters, 135, 361-368.

[5] Stoll, S., Camgoz, N. C., Hadfield, S., & Bowden, R. (2020). Text2Sign: Towards Sign Language Production Using Neural Machine Translation and Generative Adversarial Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1917-1926.

[6] **Camgoz, N. C., Koller, O., Hadfield, S., & Bowden, R.** (2018). Neural Sign Language Translation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 7784-7793.

[7] **Simonyan, K., Vinyals, O., Graves, & Kavukcuoglu, K.** (2016). WaveNet: A Generative Model for Raw Audio. arXiv preprint arXiv:1609.03499.

[8] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

[9] **Molchanov, P., Yang, X., Gupta, S., Kim, K., Tyree, S., & Kautz, J.** (2015). Hand gesture recognition with 3D convolutional neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 1–7.