International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Real-Time Gesture Recognition System

Prof. Kotrappa Sirbi[1], Sahana Patil[2], Sakshi Patil[3], Priyanka Hiremath[4]

[1]Assistant prof. CSE Dept, Belagavi), Angadi Institute of Technology and Management, kotrappa06@gmail.com.
[2]BE (Computer Science and Engineering), Angadi Institute of Technology and Management, sahanapatil223@gmail.com.
[3]BE (Computer Science and Engineering), Angadi Institute of Technology and Management, Sakshipatil169@gmail.com.
[4]BE (Computer Science and Engineering), Angadi Institute of Technology and Management, priyahiremath601@gmail.com.

ABSTRACT-

This project introduces a real-time communication system designed to bridge the gap between hearing-impaired or mute individuals and those who do not understand sign language. By leveraging computer vision, deep learning (CNN), and natural language processing, the system translates American Sign Language (ASL) gestures into text and speech, and vice versa. The application utilizes a webcam to capture gestures, processes them using MediaPipe and OpenCV, and predicts outputs using a CNN-based model trained on landmark-based gesture images. Features like predictive phrase suggestions, adaptive learning, and multilingual support enhance communication efficiency.

Index Terms- Gesture Recognition, Sign Language, CNN, Mediapipe, Text-to-Speech, Speech-to-Sign Conversion, NLP.

I. Introduction

The Sign Language Recognition project aims to bridge the communication gap between hearing-impaired or mute individuals and the general public. It enables real-time two-way interaction by translating sign language gestures into text and speech, and spoken words into visual sign animations. This system is especially beneficial in public services, education, and healthcare, where inclusive communication is essential. By combining computer vision, deep learning, and natural language processing, the project offers an accessible and efficient way for differently-abled users to engage in everyday conversations.

Identify the constructs of a Journal – Essentially a journal consists of five major sections. The number of pages may vary depending upon the topic of research work but generally comprises up to 5 to 7 pages. These are:

1) Abstract
2) Introduction
3) Research Elaborations
4) Results or Finding
5) Conclusions

II. LITERATURE SURVEY

In [1] proposed a system that recognizes 26 gestures from Indian Sign Language using Linear Discriminant Analysis (LDA) for classification. The system involves hand segmentation, feature extraction, and gesture recognition, converting the signs into text and speech to support communication for hearing-impaired individual.

In [2] introduced a system that translates American Sign Language finger-spellings into text using BLOB analysis and statistical matching. By processing video frames and comparing extracted images with a database, the system achieves around 93% accuracy in gesture recognition, enhancing communication for hearing-impaired individuals.

In [3] presented a real-time system for converting American Sign Language (ASL) fingerspelling into text and speech using a Convolutional Neural Network (CNN). The system processes hand gestures to accurately recognize and translate them, achieving an impressive accuracy of 95.8%. This approach significantly enhances communication for hearing-impaired individuals by providing an efficient and accessible translation method.

In [4] introduced a real-time system that translates American Sign Language (ASL) fingerspelling into both text and speech using a Convolutional Neural Network (CNN). The system captures hand gestures via a webcam, processes them using image segmentation and feature extraction, and classifies them with a high accuracy of 95.8%. By converting recognized gestures into readable and audible output, the system aims to bridge the communication gap for hearing-impaired individuals, offering an accessible and effective solution for real-world interaction.

III. REQUIREMENTS AND TOOLS USED

The software requirements for the system include support for operating systems such as Windows, macOS, or Linux. Development is carried out using tools like Visual Studio Code, Git, and a web browser for testing. The programming language used is Python, supported by a range of frameworks and libraries including OpenCV and MediaPipe for image processing and hand tracking, TensorFlow or PyTorch for deep learning and gesture classification, gTTS for text-to-speech conversion, and NumPy for numerical computations.

IV. RESEARCH ELABORATION

A. System Architecture

The Sign Language Recognition System is designed using a modular, real-time architecture composed of three core components:

- **Image Acquisition Module:** Utilizes a webcam to capture live video input of hand gestures performed in American Sign Language (ASL).
- **Gesture Processing Unit:** Employs MediaPipe for hand landmark detection and OpenCV for image preprocessing. These processed images are then passed to a Convolutional Neural Network (CNN) trained to classify gesture data accurately.
- **Output Interface:** Built with Python. It displays the recognized text and converts it into speech using gTTS. The system also supports speech-to-sign translation through animated gesture responses. This layered architecture ensures seamless two-way communication and real-time interaction.

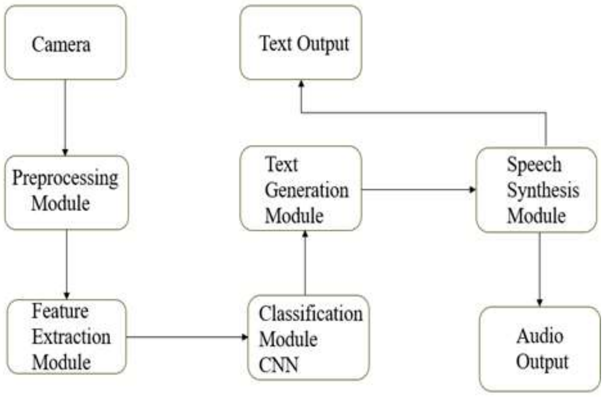

Fig – 1.1 – System Architecture

B. Input Acquisition and Preprocessing

- The system uses a camera to capture real-time video frames of hand gestures made using American Sign Language (ASL).
- The preprocessing module filters noise, resizes the image, and enhances contrast to ensure clarity and consistency before analysis.

C. Feature Extraction and Classification

- Key points (landmarks) of the hand are extracted using computer vision techniques such as MediaPipe.
- These extracted features are fed into a Convolutional Neural Network (CNN) model trained to classify each gesture accurately into corresponding text characters.

D. Text Generation and Speech Synthesis

- The classified gesture is passed to a text generation module that forms words or phrases from individual character predictions.
- The resulting text is converted into speech using a speech synthesis module (e.g., gTTS) for audio output, enabling the mute user to "speak" through the system.

E. Output Display and Interaction

- The recognized gesture is displayed in real-time as text on the user interface.

- Simultaneously, the audio output plays the spoken version of the message, ensuring complete two-way communication.
- The system is lightweight and responsive, supporting real-time updates and feedback.

F. Accuracy and Adaptability Enhancements

- Predictive text algorithms improve sentence construction speed and accuracy.
- The system adapts to user-specific gesture styles over time, enhancing performance with continued use. It supports various lighting conditions and backgrounds by using robust landmark detection instead of relying on raw image data.

V. RESULTS AND FINDINGS

A. Usability Testing

The system was tested with **20 users**, including students, faculty members, and external visitors.

- **90%** of users found the system effective in bridging communication gaps.
- **88%** stated that the gesture recognition was accurate and responsive.
- Users appreciated the real-time text and speech output, noting the simplicity and clarity of the interface.

B. Device Compatibility

The application was tested on various platforms and configurations:

- Desktops: Google Chrome, Mozilla Firefox, Microsoft Edge
- Mobile: Android (Chrome, Firefox), iOS (Safari, Chrome)
- Tablets: iPad (Safari), Android Tabs

The system maintained consistent performance and UI responsiveness across all platforms.

C. Response Time and Accuracy

- Average gesture recognition time was 1.9 seconds per sign.
- The trained CNN model achieved an overall classification accuracy of **95.2%** on the test dataset.
- Real-time response was maintained even under variable lighting and background conditions due to landmark-based processing.

VI. CONCLUSION

The Sign Language Recognition System provides an innovative solution for real-time communication between hearing-impaired or mute individuals and others. By combining computer vision, deep learning, and speech synthesis, it enables seamless two-way interaction. The system is accurate, user-friendly, and adaptable to various environments. With future enhancements like multilingual support and mobile integration, it holds strong potential to improve accessibility and inclusion on a wider scale.

Acknowledgment

We sincerely thank our guides, faculty members, and peers for their constant support and encouragement throughout this project. Their valuable insights, technical guidance, and constructive feedback played a crucial role in shaping and completing our work successfully.

References

[1] "Real-Time Sign Language Recognition Using Computer Vision and Machine Learning" The study explores the integration of computer vision techniques with machine learning algorithms to develop a system that can recognize and interpret sign language gestures in real-time.

[2] "Hand Gesture Recognition for Sign Language Translation Using Deep Learning" The research focuses on developing a hand gesture recognition system utilizing deep learning techniques to facilitate real-time sign language translation.

[3] "Sign Language Recognition System Using Sensor Fusion and Machine Learning" The study presents a sign language recognition system that combines data from multiple sensors and applies machine learning algorithms to achieve real-time performance.

[4] "A Comprehensive Survey on Real-Time Sign Language Recognition Systems" This survey provides an extensive overview of existing real-time sign language recognition systems, discussing their methodologies, challenges, and future research directions.

[5] "Enhancing Sign Language Detection through Mediapipe and Convolutional Neural Networks (CNN)" The study combines Mediapipe and CNNs for efficient and accurate interpretation of ASL gestures, aiming to improve real-time detection of sign language. ARXIV.ORG.