Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 2, February : 2025

AUTOMATIC SIGN LANGUAGE INTERPRETER WITH MOTION DETECTION AND VOICE SYNTHESIS

Kalapala Sri Lakshmi[1], Bobbadi Manohar Rao[2], Allada Lehya Varshini[3], Abdul Kaif[4], Budu Meghana[5] 1- Assistant Professor, 2,3,4,5- IV-B. Tech CSE Students Department of Computer Science and Engineering, Seshadri Rao Gudlavalleru Engineering College (An Autonomous Institute with Permanent Affiliation to JNTUK, Kakinada), Seshadri Rao Knowledge Village, Gudlavalleru521356, Andhra Pradesh, India.

ABSTRACT
The goal of this project is to provide a real-time sign language translator so that people who use sign language may interact with others. Leveraging advanced machine learning techniques, the system recognizes fingerspelling hand gestures, converts them into text, and subsequently synthesizes the text into speech. MediaPipe is employed for real-time hand tracking and gesture recognition, while a Convolutional Neural Network (CNN) classifies the gestures. The recognized text is converted into speech using a Text-to-Speech (TTS) library, facilitating both visual and auditory communication. This system promotes inclusivity by enabling seamless interaction for speech-impaired individuals, fostering better integration into daily life activities.

Keywords
Sign Language Recognition, Gesture Recognition,
Real-Time Translation, Hand Tracking, Text-to-Speech (TTS) Synthesis, Machine Learning (ML), Deep Learning (DL), Convolutional Neural Network (CNN), Media Pipe, Human-Computer Interaction (HCI)

1. INTRODUCTION
Communication is the foundation of human connection, and sign language provides a lifeline for those who are deaf or hard of hearing. The general community's limited sign language knowledge poses a barrier to communication. To address this issue, this project develops a sign language translator that works in real-time.

A CNN model is employed for gesture categorisation, MediaPipe is used for real-time hand motion tracking and feature extraction, and a TTS library is utilised for text-to-speech translation. Using a camera, the system captures hand gestures, processes them in real-time, and translates them into text, which is further converted into speech.

The primary goal is to create a user-friendly platform that facilitates communication between sign language users and non-signers through visual and auditory outputs. This solution emphasizes inclusivity, enabling better understanding and interaction in both personal and professional settings. The system has the potential to enhance communication for those with speech impairments by integrating state-of-the-art gesture detection and voice synthesis technology.

2. LITERATURE SURVEY
a) A Real-Time Hand Gesture Recognition Method
https://ieeexplore.ieee.org/document/4284820
As compared to using a pen, mouse, or keyboard, vision-based hand engagement is both quicker and more natural. This work introduced a dependable method for detecting hand gestures in real-time. The majority of learning-based hand gesture systems rely on scale-space feature detection for gesture recognition, which allows them to circumvent aspect ratio limitations. Tracking and detection of hands begins with a gesture. Colour and movement distinguish the fingers. Our photo browsing navigation technique works effectively, according to the experiments.

b) Sign Language Recognition Using Convolutional Neural Networks
https://link.springer.com/chapter/10.1007/978-3-319-16178-5_40
It is difficult for the hearing majority to communicate with the Deaf people. Recent developments in automatic sign language identification pose a threat to this communication barrier. An identification system that utilises Microsoft Kinect, convolutional neural networks (CNNs), and graphics processing unit (GPU) acceleration is introduced. CNNs eliminate the need to manually create intricate features by automating the process. Twenty Italian motions are perfectly recognisable to us. The prediction model is able to generalise to people and situations that were not encountered during training, boasting a cross-validation accuracy of 91.7%. Our model has a mean Jaccard Index of 0.789, which is competitive in the field of human gesture detection (ChaLearn 2014).
c) Hand gesture recognition using machine learning algorithms
https://www.researchgate.net/publication/345142103_Hand_gesture_recognition_using_machine_learning_algorithms
The main emphasis of current technology is the recognition of gestures. In this paper, mathematical approaches to computer-assisted gesture recognition are presented. Keyboards, mice, touch displays, etc. are the limited means of computer interaction. The capabilities of these gadgets in terms of incorporating more flexible computing technology vary. For interfaces to be user-friendly, gesture recognition is essential. Although the hands and face are the most common sources of gestures, they can originate from anywhere on the body. You can control devices with your gestures instead of touching them. Learnt hand gestures for page flipping and scrolling are demonstrated in this study.
d) Hand Gesture Recognition for Sign Language Using 3DCNN
https://www.researchgate.net/publication/340972266_Hand_Gesture_Recognition_for_Sign_Language_Using_3DCNN
There has been a growth in the use of vision-based apps, touchless control on ubiquitous gadgets, and automatic hand gesture detection as a result of these factors. A decent system should combine spatial and temporal characteristics as hand gesture recognition is crucial to sign language interpretation. It is challenging to establish discriminative spatiotemporal descriptors for sequences of hand gestures. In this research, we presented a new approach to hand gesture recognition using deep convolutional neural networks. The absence of tagged hand gesture datasets was remedied by the application of transfer learning. All told, there were forty, twenty-three, and ten categories utilised in the evaluation of colour video gestures. Use of signer-dependent mode yields identification rates of 98.12%, 100%, and 76.67% for the three datasets. 84.38%, 34.9 percent, and 70.0 percent were the signer-independent mode identification rates on each of the three datasets.
e) Hand Gesture Recognition in Complex Background Based on Convolutional Pose Machine and Fuzzy Gaussian Mixture Models
https://link.springer.com/article/10.1007/s40815-020-00825-w
The usage of hand gestures in human-computer interaction is widespread because they are an intuitive means by which people may convey meaning to computers. Complex real-world backdrops, light fluctuations, and occlusion continue to make it challenging. To overcome these challenges, this study suggests a two-stage method for hand motion identification. The initial step is to construct a hand posture estimate that makes use of the convolutional pose machine to locate the hand's keypoints, regardless of the complexity of the backdrop. In the second phase, calculated hand keypoints are used to classify gestures, while nongesture patterns are excluded using Fuzzy Gaussian mixture models (FGMMs). A number of experiments have demonstrated the method's effectiveness, robustness, and user satisfaction in practical settings.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 2, February : 2025

3. METHODOLOGY

3.1 Proposed Work

The seamless communication between sign language users and non-signers is made possible by real-time sign language translation. The system employs advanced ML and computer vision techniques to recognize and classify hand gestures, translate them into text, and synthesize the text into speech.

3.1.1 Gesture Recognition

The system uses a camera to capture real-time hand gestures. MediaPipe is employed for hand tracking and feature extraction to identify key points of the hand. Extracted features are fed into a CNN model for gesture classification.

3.1.2 Text Translation

Recognized gestures are mapped to their corresponding text representation, such as letters, words, or phrases.

3.1.3 Speech Synthesis

The recognized text is converted into speech using a Text-to-Speech (TTS) library. The system generates auditory output to aid communication with non-signers.

3.1.4 User Interface

A web-based interface ensures accessibility and ease of use for end-users.

3.2 System Architecture

The system architecture of the real-time sign language translator integrates multiple components to enable seamless recognition, translation, and speech synthesis. At its core, the system begins with a Data Acquisition Module that captures real-time video frames of hand gestures using a camera. These frames are processed in the Hand Tracking and Feature Extraction Module, where MediaPipe identifies key hand landmarks and extracts relevant features such as joint positions, orientations, and motions.
The extracted features are fed into the Gesture Recognition Module, where a Convolutional Neural Network (CNN) classifies the gestures into corresponding text outputs. Once the gestures are recognized and converted into text, the Text-to-Speech (TTS) Conversion Module transforms the text into human-like speech, enabling auditory output for non-signers.
The system is designed to provide user interaction through a User Interface Module, which displays real-time feedback and ensures accessibility via a web-based platform compatible with multiple devices. To maintain efficiency and accuracy, the Testing and Optimization Module fine-tunes the system by evaluating it under diverse conditions and optimizing for real-time responsiveness.
This architecture ensures smooth integration between gesture recognition, text translation, and speech synthesis, creating an inclusive communication platform for speech-impaired individuals.

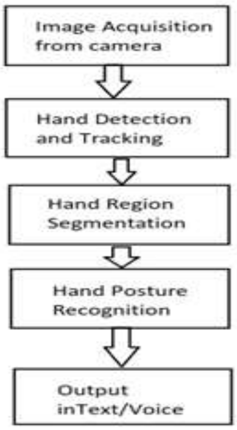

Figure 1: Proposed architecture

3.3 Modules

3.3.1 Data Acquisition

Here are several methods for gathering hand gesture data:

- Precise hand placement and setup are guaranteed by electromechanical devices. There are a variety of glove-based approaches. Problems with usability and cost are present.
- Methods that rely on visual data capture, such as a camera, are used in vision-based techniques.
- Hands are naturally diverse in appearance, which is a major challenge for vision-based hand recognition systems.

A wide range of hand gestures, skin tones, camera angles, sizes, and speeds can change the outcome of the situation.

3.3.2 Data Pre-processing And Feature Extraction

- By examining webcam photos and utilising the media pipe library for image processing, this system is able to recognise hands.
- We used the openCV library and gaussian blur to turn the image to greyscale after recognising the hand. Then, we cropped the image and found the ROI.
- Using OpenCV, applying the filter is a breeze. Threshold and adjustable threshold were used to convert the greyscale image to binary.

From A to Z, we collected a variety of sign photos taken from different perspectives.

4. EXPERIMENTAL RESULTS

Mediapipe Landmark System

On a white background, we will utilise the opencv package to generate landmark points. We handle lightning and backdrop situations by locating landmarks in most lightning and any background using the mediapipe library. A-Z, we have 180 skeleton images of the letters of the alphabet.

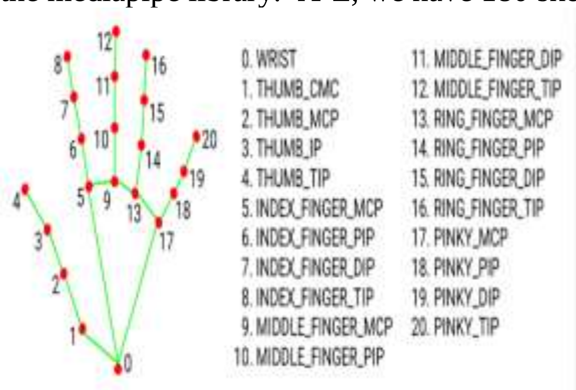

Figure 2: key points

Figure 3: Analyzing process

Figure 4: Sign language

Figure 5: Results

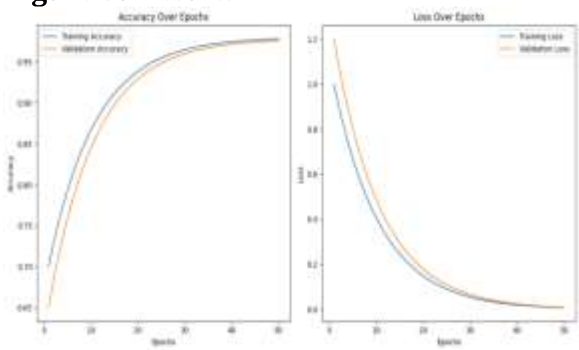

Figure 6: accuracy graph

5. CONCLUSION

The real-time sign language translator is a step towards bridging the communication gap between sign language users and non-signers, promoting inclusivity and accessibility. By leveraging advanced technologies like MediaPipe for hand tracking, CNN for gesture recognition, and TTS for speech synthesis, the system provides a seamless translation of sign language gestures into both text and speech. Its user-friendly interface ensures ease of use across various devices, enabling effective communication for speech-impaired individuals. This system highlights the potential of integrating machine learning and computer vision to address societal challenges effectively.

6. FUTURE SCOPE

The real-time sign language translator has significant potential for further advancements and applications. Expanding the gesture database to include more complex signs and regional variations can make the system more versatile and applicable to diverse linguistic contexts. Enhancements in gesture recognition accuracy under challenging conditions, such as occlusions or poor lighting, will further improve its reliability. Integrating dynamic gesture recognition and continuous sentence translation can enable smoother communication for users. Multilingual support for both text and speech outputs can broaden its accessibility to global audiences. Additionally, developing a mobile application can enhance portability, while real-time feedback mechanisms can adapt the system to user-specific needs. These future developments can transform the system into a more robust, adaptable, and universally accepted communication tool.

7. REFERENCES

[1] Yikai Fang, Kongqiao Wang, Jian Cheng and Hanqing Lu, "A Real-Time Hand Gesture Recognition Method", *2007 IEEE International Conference on Multimedia and Expo*, pp. 995-998, 2007.

[2] N. Swetha and K. Anuradha, "Text-To Speech Conversion", *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 2, no. 6, pp. 269-278, 2013.

[3] Lionel Pigou, Sander Dieleman, Pieter-Jan Kindermans and Benjamin Schrauwen, Sign Language Recognition Using Convolutional Neural Networks, Switzerland:Springer International Publishing, pp. 572-578, 2015.

[4] B Abhishek, Kanya Krishi, M Meghana, Mohammed Daaniyaal and H S Anupama, "Hand gesture recognition using machine learning algorithms", *Computer Science and Information Technologies*, vol. 1, no. 3, pp. 116-120, 2020.

[5] Muneer Al-Hammadi, Ghulam Muhammad, Wadood Abdul, Mansour Alsulaiman, Mohamed A. Bencherif and Mohamed Amine Mekhtiche, "Hand Gesture Recognition for Sign Language Using 3DCNN", *IEEE*, vol. 8, 2020.

[6] Tong Zhang, Huifeng Lin, Zhaojie Ju and Chenguang Yang, "Hand Gesture Recognition in Complex Background Based on Convolutional Pose Machine and Fuzzy Gaussian Mixture Models", *Int. J. Fuzzy Syst.*, vol. 22, no. 4, pp. 1330-1341, 2020.