INDIAN SIGN LANGUAGE LEXICAL INTERPRETER

Project Reference No.: *47S_BE_5019*

College *: S.D.M. College of Engineering and Technology, Dharwad*
Branch *: Department of Information Science and Engineering*
Guide(s) *: Dr. Pushpalatha S. Nikkam*
Prof. Varsha S. Jadhav
Student (S) *:Ms. Spoorti S. Aldi*
Mr. Omkar Belamkar
Mr. Suraj Bhajantri
Mr. Vishal Nagmule

Keywords

Indian Sign Language, gestures, Long Short-Term Memory, OpenCV, Deep Learning.

Abstract

Sign language is a language that comes out when deaf and dumb people could not communicate with the normal people. However, normal people have little knowledge about sign language used by deaf and dumb people to communicate. We have extracted a total of 40 articles from well-known online databases using selected keywords. The review shows that the vision-based hand gesture recognition research is an active field of research, with many studies conducted, resulting in dozens of articles published annually in journals and conference proceedings. Most of the articles focus on three critical aspects of the vision-based hand gesture recognition system, namely: data acquisition, data environment, and hand gesture representation. We have also reviewed the performance of the vision-based hand gesture recognition system in terms of recognition accuracy. For the signer dependent, the recognition of alphabets is done in wide range. The lack in the progress of gesture recognition of words used in sign language could indicate that more work is needed towards a practical vision-based gesture recognition system. Many researches have been done to make communication possible between these people using various algorithm like SVM, KNN, CNN, RNN, PCA, HOG and so on. In this paper we have proposed a method to recognize the hand gestures using CNN algorithm and MediaPipe open-source framework. Another feature we added is converting the interpreted text to speech.

Introduction

Gestures can be categorized into hand and arm gestures (recognition of hand poses, sign languages, and entertainment applications), head and face gestures (such as nodding or

shaking of the head, the direction of eye gaze, opening the mouth to speak, winking, and so on), and body gestures (involvement of full-body motion).

The task is accomplished using Image processing and Machine learning. Machine-learning and Image processing are very powerful tools often used for image classification and recognition. Image Processing deals with the image, its properties, and the operations performed on it for getting some information from the images. Machine learning is the study of algorithms and statistical data used to perform tasks using various data patterns and inferences. In this project, the collection of the images of the sign language are to be done using a camera. The images are then processed and the features are extracted using image processing. These images are them compared from the available datasets and by implementing deep learning, the signs are interpreted. The data is displayed on a display that helps the person in front of a deaf/mute person understand the sign language We have developed a simple and lightweight deep learning algorithm which can detect static sign-language gestures.

The Indian Sign Language is predominantly used in the South Asian region. It has several variants depending on the region as well. The ISL is still evolving and works are being done to standardize it by ISLRTC. Indian Sign Language Research and training Center (ISLRTC) is an autonomous organization under the administrative and financial control of Department of Empowerment of Persons with Disabilities Ministry of Social Justice & Empowerment, Government of India.

It appears that the past works reviewed by researchers had left a gap; they had not examined the challenges and future direction of the vision-based hand gesture recognition system. Based on this, the current paper will address this gap by reviewing existing literature to identify the progress of research in gesture-based recognition systems of words for the present and for future directions.

Problem Statement

There has always been a communication barrier between deaf and mute people and the speaking community. This is especially apparent during the time of an emergency. Most of the times a human translator is employed for the translation. But everyone cannot afford a human translator and during emergencies, the availability of a human translator cannot be assured. This kind of communication not only makes the disabled person dependent on such professionals but also rather expensive in addition to not being practical for day-to-day communicative requirements of the people. A replacement of such professionals by a software tool for automatic recognition and translation of a sign language into text and then

to a vocal language is needed in helping hearing impaired people to communicate with the society and also to help them to be independent up to certain level. Such an automated tool would acquire gestures, analyze, recognize and then produce equivalent sentences in text and then in a vocal language.

Objective of the Proposed Work

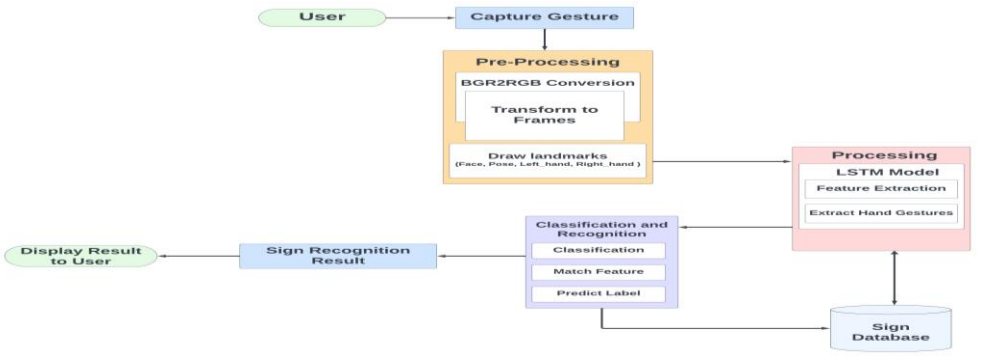
Sign languages, a visual-gestural mode of communication, have emerged as crucial tools for bridging this communication gap. In India, where linguistic diversity is a hallmark, Indian Sign Language (ISL) serves as a vital means of communication for the Deaf and Hard of Hearing (DHH) community. Some of the main objectives for developing ISL word interpreter are:

- To develop a system that can automatically recognize and translate words from Indian Sign Language (ISL) to text or speech and under diverse conditions, including variations in signing speed, sign size, and hand shapes.
- To provide a communication tool for hearing and speech impaired people in India who use ISL as their primary language.
- To design a system capable of real-time recognition, allowing for seamless and immediate communication between ISL users and those unfamiliar with the language.
- To create an intuitive user interface that is accessible to both ISL speakers and non-speakers, promoting inclusivity and ease of use.

Proposed Methodology

While sign language, which is the major communication tool for many deaf people, provides a solution, universal understanding remains a challenge. Currently, learning sign language and locating trained sign language interpreters are substantial challenges. This technology aims to empower people with speech or hearing impairments to communicate using their preferred sign language, so increasing inclusivity and accessibility.

Fig. Methodology



The methodology has several steps to recognize the gesture at the end:

1. It starts with the appearance of real time video and then captures gesture as input which will be validated for landmarks and is captured and sent to next phase. This is a key phase in the sign recognition process. Camera interfacing is important for capturing photos via a webcam. Many laptops now include built-in cameras, making it easier to capture and handle photographs. The embedded camera detects hand movements and positions.

2. The captured real time video is converted to BGR to RGB and the transformed to frame and the video stream is live and cannot be changed and will display the label predicted at the end. OpenCV, a computer vision library in Python, is widely used for image analysis, processing, detection, and identification. Image processing techniques such as image enhancement, segmentation, and feature extraction could then be used to extract relevant information about the hand gestures as shown .

3. For extracting the features points, MediaPipe enables multi-modal machine learning pipelines for computer vision applications. It provides excellent ready-to-use ML solutions. The MediaPipe Holistic pipeline integrates increased face, hand, and posture components. This pipeline enhances the model's ability to recognise facial traits as well as hand and body locations. MediaPipe aims to recognize faces and hands, extract key elements, and input them into computer vision models.

4. The landmarks in the image can be identified using the holistic model predictions in the result variables. The holistic model analyses the image and creates landmarks to identify the user's face, left and right hands, and pose. The model outputs an array. It will collect crucial point values for testing and training. We will build our datasets based on the outcomes of the previous phase. There will be 30 films for each action, each consisting of 30 frames. NumPy arrays will hold all of the frame data.

5. The LSTM can handle an entire sequence of data (e.g., a time series) without analyzing each point individually, providing valuable insights. LSTMs excel at processing various data streams, including text, audio, and time series. After collecting data, we build and train the LSTM model as in Fig. 1 to detect action across multiple frames. when the extracted features and landmarks are given to it applies the RNN algorithm and use it to match feature and predict label and send the result.

6. Then we match the feature and translate to gesture, here it compares the extracted feature points with a database of known ISL gestures to find potential matches. Then it Converts the recognized gesture into its corresponding ISL word or phrase.

7. The system will predict the label, that is it will assign a specific ISL word or phrase to the identified gesture, representing its intended communication. Then the system communicates the interpreted meaning of the captured gesture back to the user, enabling understanding and interaction in the form of text and speech.

Results

There are several test cases that can be consider but the main test cases that are necessary to run are three test cases as below. The first says an ideal image where we can observe only camera is detecting person's face, but sign and caption are not generated as in Fig. 5. The second case shows the person giving sign which is not trained so the landmarks are drawn in Fig. 6, but no sign and caption is generated. The last and main test case is the person will action the trained sign so system will detect the sign and generate caption for the same as shown like Fig. 7 shows teacher sign is detected, Fig. 8 shows deaf and Fig. 9 shows man sign detected.

Result Case I – Ideal Case

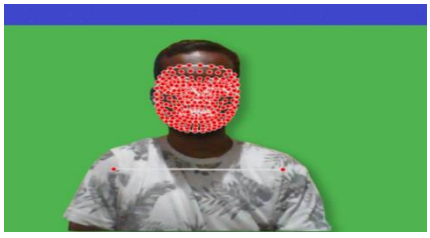


Fig 1. Ideal image

In the ideal case, we observe that the person is not performing any gesture or sign while facing towards the camera. What expected result is should be. There should be no sign and no caption will appear. Science. In the ideal state there is no action perform. There should be no. Caption generated. The blue line in the image is the. Space provided for the captions

which are generated after performing the sign. Since there is No sign or gesture Performed. There is an empty space.

Result CASE II – Detection of not Trained Sign

Fig 2. Detection of Unknow Sign

In the case i.e. Fig. 5 shows of unknown sign detection. The program identifies the landmarks and the pose performed by the user. But since there is no sign detected in the previous trained model. It does not recognize this sign.

Result CASE III – Detection of Trained Sign

In this case of detection of train sign when the user performs. The sign which is already trained to the data module. It generates caption Following to the sign recognised. In the first image, we can observe that the user has performed an Indian Sign Language for teacher. And data model recognises the sign and provide the following caption as “Teacher”. The second We observe that the user performs Indian Sign language for deaf. And the model recognises sign by providing the captions of the recognised sign as “Deaf” as shown in Fig. 6,7,8. Same As for the 3rd image, the user performs an Indian Sign language for Man. And the data model recognises the sign as “Man”. The recognition of the sign language works very well only if There is a good Intensity of light on the face of the user who is providing the sign. Presently, Using LSTM Model we can only train 3 signs. Since it requires a I and. Computing devices to train more than 3 sign languages. To perform and train models.

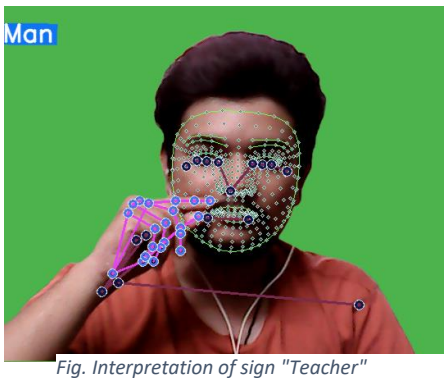


Fig. Interpretation of sign "Teacher"

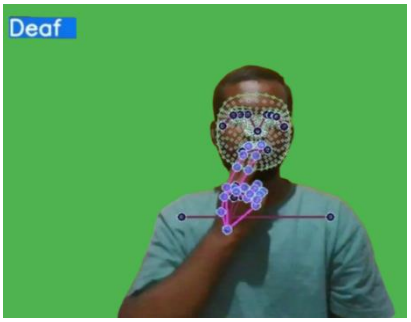


Fig. Interpretation of sign "Deaf"

Fig. Interpretation of sign "Man"

Conclusion

The system for recognizing real-time Indian Sign Language (ISL) portrays an impressive role in enhancing casual communication among people with hearing disabilities and normal persons. Sign Language Recognition System has been developed from classifying only static signs and alphabets, to the system that can successfully recognize dynamic movements that comes in continuous sequences of images. Researcher nowadays are paying more attention to make a large vocabulary for sign language recognition systems. Many researchers are developing their Sign Language Recognition System by using small vocabulary and self-made database. Large database build for Sign Language Recognition System is still not available for some of the country that involved in developing Sign Language Recognition System. The classification method of identifying the sign language is also varied from researchers. Using their own ideas and limitations for the Sign Language Recognition System, the comparison of one method to another method is still subjective. Variation of sign language in most of the country is based on their grammar and their way to present each word, such as presenting the language by word or by sentence. This paper thus concludes that to make the vision-based gesture recognition system ready for real-life application, more attention needs to be focused on the uncontrolled environment setting as

it can provide researchers the opportunity to improve the ability of the system in recognizing hand gestures in any form of environment.

Innovation of the Project

- **Existing:**

1. Human Translator: While human interpreters provide a valuable service, they can be expensive, limited in availability, and may not be fluent in all ISL dialects.

2. Generic Sign Language Tools: Tools designed for other sign languages might not accurately translate ISL due to grammatical and vocabulary differences.

3. ISL Alphabets: There are systems which are implemented to recognize only alphabets from A-Z.

- **Innovation:**

1. Interface: Will reduce the use of human translator using our interface.

2. Use of LSTM: The algorithm here we used is LSTM which recognizes the signs accurately and quickly.

3. ISL Words Interpreter: Our main innovation is to interpret the ISL words so that communication is easier.

Future Scope

- **Enhanced Gesture Recognition:** Train on broader sign variations and leverage deep learning for superior sign recognition accuracy. Training on a wider range of data encompassing regional variations, signing styles, and backgrounds can improve recognition across different scenarios.
- **Personalization and Customization:** Adapt to individual signing styles and translate regional ISL dialects for wider user base. The interpreter could be designed to recognize and translate regional variations of ISL, catering to a wider audience across India.
- **Integration with Wearable and IoT Devices:** Facilitate hands-free sign language interpretation through wearable tech like smartwatches. . Imagine a scenario where a Deaf user wearing a smartwatch can have a real-time sign language conversation, with the watch translating the signs into spoken language for the hearing person and vice versa.

- **Wider Accessibility:** Develop mobile apps and enable offline functionality for improved access to the interpreter. Equipping the interpreter with the ability to function even without an internet connection would be crucial for situations where internet access might be limited.
- **Sequential Sign Construction:** Utilize NLP and contextual awareness to understand the order and meaning of signed sentences. By incorporating contextual information from the surrounding environment or conversation history, the system can improve its understanding of the meaning conveyed through signs.