NLP-based Sign Gesture Identification for Disabled People

Pankaj Saraswat

SOEIT, Sanskriti University, Mathura, Uttar Pradesh, India

Correspondence should be addressed to Panakaj Saraswat; pankajsaraswat.cse@sanskriti.edu.in

ABSTRACT- There are many methods for identifying signs, each of which generates a word for each one. It focuses on converting sign language into an appropriate English sentence. NLP techniques are also used in addition to sign recognition. The input is a framed and split video of sign language. This booklet teaches deaf and mute people sign language. It's tough for non-blind persons to engage with blind people due to communication difficulties. To address this issue, the article suggests and describes an effective method. Language technology methods such as POS tagging and the LALR parser are used to convert identified sign words into English phrases. A number of applications are currently on the industry that allow blind people to interact with the world. Combining technology will not be able to address the problem of mobile sign language translation in daily activities. A video interpreter can assist deaf or hearing-impaired people in a variety of situations. People with hearing impairments will be able to learn sign language and have films translated into sign language as a consequence of this research. The present work may be used as a communication interface for both speech-impaired and non-speech-impaired individuals. It will assist bridge the communication gap between speech-impaired people and the rest of the population by capturing and analyzing signals, as well as recognizing and displaying output in the form of comprehensible phrases.

KEYWORDS- Communication, Hearing and speech, NLP, Parsing, Sign Language.

I. INTRODUCTION

Deaf or dumb people have a variety of difficulties when it comes to interacting with others in their everyday lives due to their hearing and speech disability. Many people with hearing and speech impairments are unable to express or utilise their skills in the outside world due to a linguistic barrier. The aim is to develop a system that will help these individuals bridge the gap between themselves and the rest of society. According to a research, India is home to 20% of the world's population with hearing and speech impairment. While hearing and speech impaired people in India communicate using Indian Sign Language, which is not widely recognized or understood by the general public, the general public uses natural language. This will need the use of an interpreter. Human interpreters are often used, although they are costly and not available to everyone[1].

The motions of the hands influence the interaction of hearing and speech impairment persons. They communicate effectively with tactile sign language, which consists of hand movements and symbols [2]. Even if it is true, hearing-impaired people have communication challenges in a culture that is largely hearing-capable. The interaction between sign language users is the subject of this research. Human language may be deciphered using natural language principles. It is a strong tool that combines language technology with artificial intelligence. NLP includes both syntax and semantics. Syntax is in charge of the ordering and grouping of words. Semantics includes word and phrase meanings, as well as compositional meanings. Lexical semantics determines the meanings of component words, whereas compositional semantics, which combines these components, determines the wider meaning. In Natural Language Processing, POS labeling is a critical step (NLP). This is where the recovery of information starts. Parsing, which deals with grammar, is another important method.

Android apps have significantly increased in capabilities, to the point that Java programs may now be executed on a mobile phone. As a result of this technology, mobile users all around the globe may now read and compose email, browse web sites, and play Java games on their phones. This trend has encouraged the use of Android apps for better communication. Hearing and speech impaired people may now use text messaging to communicate with hearing and speech impaired people. But only for conversations between callers who are deaf or hard of hearing. It's a program that will repeat whatever we say in a high-pitched voice in any image movement you choose. You may utilize this software without dialing a phone number.

Normal individuals utilize natural language to speak and engage with each other, while hearing and speech impaired persons use tactile sign language to communicate. The fierce competition in every area makes it harder for persons with impairments to compete. hearing and speech

impairmentindividuals will be helped by developing an application that will allow them to communicate effectively with a normal person [3]. Surveys show that India has almost 2.4 million hearing and speech impairment, or around 20 percent of the world's total.An interpreter is required to facilitate communication between a normal person and a hearing and speech impairment.

The motions of the hands influence the interaction of hearing and speech impairment. They communicate effectively with tactile sign language, which consists of hand movements and symbols. In a culture that is largely hearing-capable, it is true that the hearing-impaired must overcome social hurdles. Sentencing for hearing impairment interaction is the subject of this study. A person who is proficient in natural language is able to decipher human language[4]. Linguistics and Artificial Intelligence are involved. A system that can transform text (words) into human language is called NLP (Natural Language Processing). 1960 was the year when NLP's POS tagging approach was introduced for the first time. In the field of language processing, it is an essential tool. It is the simplest and most stable step in many NLP applications. It is the first stage in machine translation, information retrieval, and other applications.

Parsing is the second most significant approach in NLP. Compilers use a process called parsing. Hearing and speech impairment video output is used to build a sentence.The Natural Language Processing engine is responsible for the sentence-building process. POS tagging and Parsing are the two major strategies used in natural language processing. During the preprocessing stage, a video of a hearing impaired person interacting is recorded shown in Figure 1. Film processing procedures including segmentation, tracking, feature extraction and classification are carried out on this video as well as the identification of enacted words that appear.
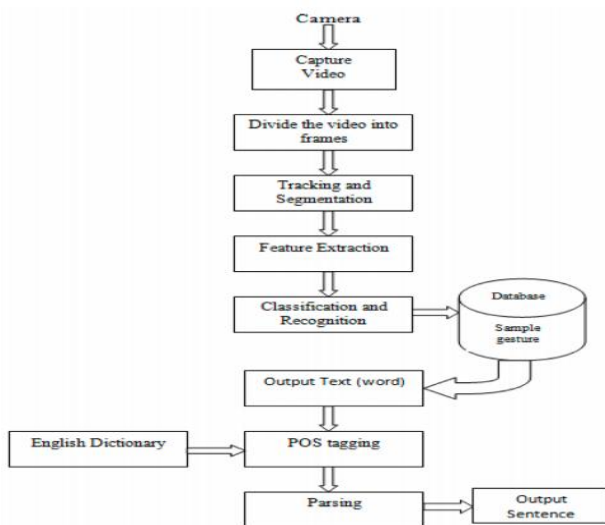

Figure 1: Systematic Flowchart of the Total Processes Involved Within the System Employed

Processing of movies before they are uploaded with the aid of a camera, a video of the hearing impaired person is taken. Using image processing methods, this footage is then preprocessed [5].
The following are the preprocessing steps:

- The framing of a picture samples
- Tracking and segmentation of data
- Extraction of a feature
- Recognition and Classification

Depending on its size and timing, video is broken down into frames. The only thing that's taken into account when shooting the video is hand movement. A 2D camera is utilized to capture the movies. The greatest curvature point is used for frame extraction. Color Saturation Value Method and Camshift Method are used to segment and track the image, respectively. With a variety of lighting situations, it produces an acceptable outcome. A pseudo 2-dimensional Hidden Markov Model (P2DHMM) is utilized to extract features. Using a HAAR-like classifier, classification and recognition are performed [6]. Using a database, the recognition process is carried out.

A. POS Tagging

Tagging is the process of assigning a description to each word obtained from the preprocessing phase in the NLP engine. There are two main steps in the process of tagging: tokenization and tagging, respectively. Tagger is the description descriptor. An interjection, verb, adjective, adverb, adverbial noun or conjunction, preposition or pronoun is indicated by the tag. i.e. Raw Text input becomes tokenized and a text is utilized to determine the matching phrase part-of-speech (POS).

B. Designing of grammar

According to grammar, lawful sentences in a given language are defined by a set of rules. These principles act in concert to create legitimate sentences, and they are interdependent. It is the simplest kind of grammar to use because it does not require any context [7]. Context-free grammar is straightforward to work with and write for others. Grammar is often damaged when rules are written in a context-free manner. When difficulties arise, these rules are modified. A multitude of techniques are employed in the creation of the grammar. The following methods can be used to analyze the use:

- Conceptual Tool:By collecting and reporting natural language, it is possible to determine the level of complexity.
- Formal Notation:formal notation is utilized by parsers to trace syntactical and semantic features of the language.

The first category focuses on the language's structure, while the second focuses on the implementation and explanation of the grammar given. Programs and programming languages have a similar connection between conceptual design and notation. Parser is a language interpreter or compiler that is

used to recognize a language's structure. The following is a representation of grammar design:

$$G = \{V, \Sigma, P, S\}$$

Where, V=Set of non-terminal

Σ= Set of terminal

P= Set of production

S= Start symbol

1) Parsing

Parsing from the bottom up is done using a Left to Right (LR) parser. As a method for parsing broad classes of context-free languages, it is flexible, shift-reduced, efficient, and helpful. With the assistance of direct computation, it is possible to establish that LR parser is a recursive program. To separate and analyze text based on formal grammar production rules, LALR parsers are simpler versions of Left to Right (LR) parsers. It takes as inputs pronouns, verbs, conjunctions, adjectives, and so on. The terms is, am, are, and article do not exist in tactile sign language. A parse tree is used to insert words and articles at the right locations in a sentence.

Image processing is used to identify gestures between hearing impaired individuals, which is superior to the previous approach and opens up the possibility of communication between hearing and speech impaired people in their daily activities utilizing sign language and video relay services, according to this research. One day, as video technology continues to develop, it may become the primary mobile communication method among those who are hard of hearing. The primary goal of this project is to build an interpreter that will translate sign language into English.hearing and speech impaired individuals' interactions will be recorded on videotape and fed into a computer program that will identify words for each actioned sign before turning it into an understandable phrase. Communication between normal and handicapped individuals is improved by this study. They can connect with hearing persons in real time via video or other technologies via a sign language interpreter, even if they're hearing impaired, hard of hearing, or speech challenged. It is believed that hearing impaired persons would learn English more effectively if they are exposed to SE (Signed English) and other parallel signing systems [8]. This is followed by a classification of the sensor data using 24 English alphabets and two punctuation marks. In the study, the topic of dynamic gestures is explored. Sensor gloves may not be able to detect dynamic motions, thus a gesture on the arm can be used to combat this [3]. The suggested approach is designed to aid persons with language impairments in sentence creation and grammatical correction.

To translate text between languages, enhanced attribute grammar is used [9]. Demonstrated a method for comparing basic grammar sentences to a set of simple images. Attribute grammar is used to demonstrate semantic analysis of natural language text in the technique described here. Android application with optimized algorithm has been developed and tested with real-time data [10]. Many additional chores, such as de-noising, de-hazing, de-rinsing & de-blurring, are also covered in detail. With our technique, the authors look at both the capability of large models and the sheer volume of data. Hearing and speech impaired individuals communicate with each other through sign language. It is followed by a model for pre-training that can handle a number of different image processing functions. Since the video is already captured and saved in the cloud, it may be interpreted by the system in a suitable manner.Their approach, on the other hand, addresses this problem to a greater extent than the method described above.

C. Data Collection

- Take video as an input. Specifications of video is as follows: Type of file: .avi, Size: 7.49mb, Length: 5 sec, Frame width: 900, Frame height: 508, Frame rate: 12 Fps as shown in Figure 2

Figure 2: Sample Frames of the Number of the Frames = Frame Time*FPS

- The following formula may be used to divide the video into frames: Framing time = number of frames*FPS as shown in Figure 2.
- For Grayscale conversion, binaries the frames before converting them to black and white as shown in Figure 3.
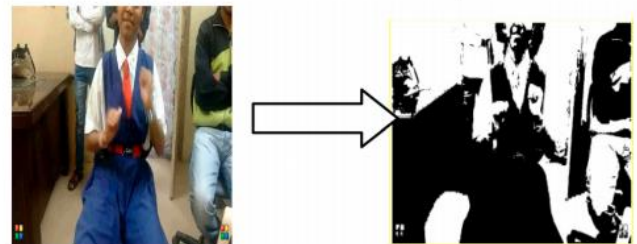

Figure 3: A Single Sample Frame from the Camshift of the Frame into Grayscale i.e. Binary Format

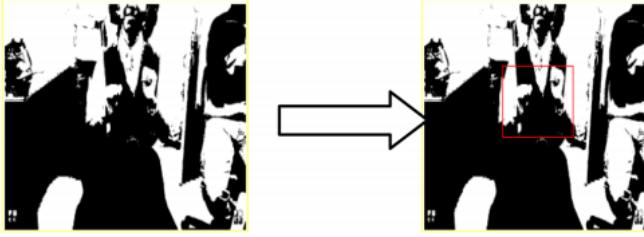- Using HSV and CAMSHIFT, segmentation and tracking are conductedas shown in Figure 4.

Figure 4: A Single Sample Frame from Grayscale to More Subjected Tracking to the Gesture

- The P2DHMM method is used to extract features from the frame as shown in Figure 5
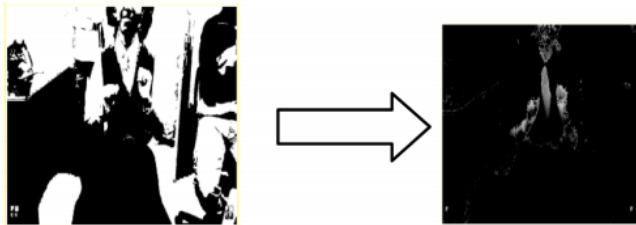

Figure 5: A Single Sample Frame from Grayscale to More Subjected Tracking for Extract Features from the Frame

- With the help of the Haar Classifier, which identifies the model feature using 33 samples of expression, the hand movement is classified and recognized.
- Next, the NLP engine, which consists of POS tagger and LALR parser, receives the output words from the NLP engine. As the words are tagged by the POS tagger, they are then parsed by the Parser to produce sentences. The following is the POS tagging and parsing of the words above: I (N), Active (Adj).

II. DISCUSSION

The hearing and speech impairment specialist evaluated the sentence received from the films for analytical purposes. Hearing and speech impairment special school teachers manually interpreted five videos that were converted to sign language using the system's output, which yielded five phrases. According to that, the output of this system and the manual interpretation are quite comparable in terms of their results. It is therefore accurate to within about 90 percent.Using this program, hearing and speech impaired people may effortlessly communicate with anybody, anywhere. Speech-to-sign transfer is also supported by this initiative.Video Relay Service and outfit-7 are two of the most important software components (VRS). A system that combines all of these components is possible. Nothing except American Sign Language (ASL) has been included in our project as a vital means of communication amongst speech impaired people (American Sign Language).

Only the right hand is used to sign all letters, with the palm towards the observer. SE (Sign English) is a good equivalent to English in terms of its manual nature and structure. If hearing impaired people are exposed to SE and other parallel signing systems, they will acquire English more quickly. A sign language interpreter (SE) utilizes two types of gestures: sign words and sign markers. Each Sign word stands for a separate entry in a Standard English dictionary. As part of our research, we are implementing the Sign Word idea, which is beneficial in converting Sign Language into words, as well as developing algorithms in Natural Language Processing (NLP) that can handle extended sentences. They are signed in the same sequence as the words appear in an English phrase, with the exception of one. ASL is the primary source of signs in SE. These signals, however, are now used in the same sequence and with the same meaning as English phrases in order to avoid confusion. In addition, hearing impaired people may use this program for mobile sign translation (VSR) as well as UTF-7, which allows them to communicate in daily activities without dialing numbers.

III. CONCLUSION

This paper's primary goal is to explain the importance of sign language interpretation and to demonstrate how hearing and speech impairment sign language may be converted into words to help them in their daily activities. A restricted number of words were used in this study, and short phrases were used to test the results. In the future, we plan to expand our work by including a big database that contains the most possible terms and to build NLP algorithms that can handle extended phrases. ISL to English conversion is not a simple word-for-word translation, as is the case with other languages. Language learners have several challenges when it comes to identifying pronouns, prepositions, and other grammar structures. In developing the application, the prior work on NLP and gesture detection would be quite helpful. The demand for fresh software and inventive methods to make something useful will never go away. So, in order to create the application, an algorithm must be written.

Detecting facial expressions might be added to the sign recognition program in the future, which would help us punctuate sentences appropriately. In this way, we can better understand how someone is feeling.With the aid of sign language trainers and hearing and speech impaired individuals, this effort can truly contribute to the construction of a sign language interpretation system if we adhere to certain linguistic rules. In order to bridge the communication gap between hearing impaired and normal individuals, assistive systems for the hearing impaired individuals are needed.

REFERENCES

[1] Basar S, Adnan A, Khan NH, Haider S. Color Image Segmentation Using K-Means Classification On RGB Histogram. Recent Adv Telecommun Informatics Educ Technol. 2014;

[2] Mohandes M, Liu J, Deriche M. A survey of image-based Arabic sign language recognition. In: 2014 IEEE

11th International Multi-Conference on Systems, Signals and Devices, SSD 2014. 2014.

[3] Wu CH, Chiu YH, Guo CS. Text generation from Taiwanese sign language using a PST-based language model for augmentative communication. IEEE Trans Neural Syst Rehabil Eng. 2004;

[4] Rajam PS, Balakrishnan G. Real time Indian Sign Language Recognition System to aid deaf-dumb people. In: International Conference on Communication Technology Proceedings, ICCT. 2011.

[5] Wilson AD, Bobick AF. Parametric hidden Markov models for gesture recognition. IEEE Trans Pattern Anal Mach Intell. 1999;

[6] Dabre K, Dholay S. Machine learning model for sign language interpretation using webcam images. In: 2014 International Conference on Circuits, Systems, Communication and Information Technology Applications, CSCITA 2014. 2014.

[7] Zimmermann M, Chappelier JC, Bunke H. Offline grammar-based recognition of handwritten sentences. IEEE Trans Pattern Anal Mach Intell. 2006;

[8] Mehdi SA, Khan YN. Sign language recognition using sensor gloves. In: ICONIP 2002 - Proceedings of the 9th International Conference on Neural Information Processing: Computational Intelligence for the E-Age. 2002.

[9] Surabhi MC. Natural language processing future. In: 2013 International Conference on Optical Imaging Sensor and Security, ICOSS 2013. 2013.

[10] Jung C, Kim C, Chae SW, Oh S. Unsupervised segmentation of overlapped nuclei using bayesian classification. IEEE Trans Biomed Eng. 2010;