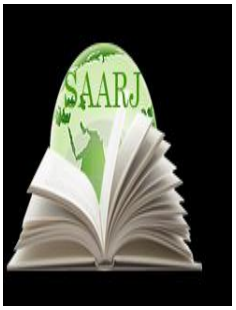
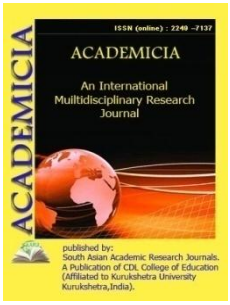


ACADEMICIA
An International Multidisciplinary Research Journal
(Double Blind Refereed & Peer Reviewed Journal)



DOI: 10.5958/2249-7137.2021.02248.5

NLP BASED SIGN GESTURES RECOGNITION SYSTEM

Mr Madhav Singh Solanki*

*SOEIT, Sanskriti University,
Mathura, Uttar Pradesh, INDIA
Email id: madhavsolanki.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 available on the market that allows 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.*

1. 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 utilize 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[1]. 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.

Hearing and speech impaired people's interactions are influenced by hand movements. Tactile sign language, which comprises of hand motions and symbols, is an efficient way for them to communicate. Even if this is true, hearing-impaired individuals have communication difficulties in a society that is mainly deafeningly[2].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[3].

Normal people interact with one another using normal language, while hearing and speech challenged people communicate through tactile sign language. In every field, intense competition makes it more difficult for people with disabilities to participate[4]. Individuals with hearing and speech impairments will benefit from the development of an application that allows them to converse successfully with a normal person. According to surveys, India has about 2.4 million people with hearing and speech impairment, accounting for around 20% of the global total. To assist communication between a normal individual and someone who has a hearing or speech disability, an interpreter is needed[5].

The interplay of hearing and speech disability is influenced by hand movements[6]. Tactile sign language, which comprises of hand motions and symbols, is an efficient way for them to communicate. It is true that the hearing-impaired must overcome social barriers in a society that is mainly deafeningly. This research focuses on sentencing for hearing impairment interactions. Human language can be deciphered by someone who is fluent in natural language. Artificial Intelligence and linguistics are both engaged. NLP refers to a system that can convert text (words) into human language (Natural Language Processing) [7]. The POS tagging method was presented for the first time in 1960, according to NLP. It is a necessary instrument in the area of language processing. In many NLP applications, it is the simplest and most stable step. In machine translation, information retrieval, and other applications, it is the initial step[8].

2.LITERATURE REVIEW

Syed Atif Mehdi et.al discussed a review on Sensing gloves for Sign Language Recognition using artificial neural networks[9]. After that, the sensor data is classified using 24 English alphabets and two punctuation marks. The subject of dynamic gestures is investigated in this research. Because sensor gloves may be unable to detect dynamic movements, an arm gesture may be employed to compensate. The main goal of the paper's approach was to see whether a sensor glove could be used to recognize various sign languages. Because sensor gloves can only

recognize a portion of sign language, as shown in the article, additional sensors are needed to detect the full form of sign language.

Chung-Hsein Wu et al. discuss a method for forming Chinese phrases from Taiwanese Sign Language for persons with hearing problems[3]. The proposed method is intended to assist people with language disabilities in constructing sentences and correcting grammatical errors. Enhanced attribute grammar is used to translate text across languages. On the basis of context-free grammar, attribute grammar formalizes the semantics of a language. This article offers efficient methods for implementing the famous Viola Jones algorithm in a real-time environment utilizing Local Binary Pattern features for hand motion recognition.

ChandhanaSurabhi M discussed about the development of natural language interfaces to interact with robots[10]. The user will be able to access apps more easily with this approach. A technique for comparing fundamental grammar phrases to a collection of simple pictures was shown. In the method presented here, attribute grammar is utilized to show semantic analysis of natural language text. Real-time data was used to build and test an Android application with an improved algorithm. This algorithm implementation has not been used to create a robust and real-time algorithm. A significant lot of research has been done to automate the process of sign language interpretation using image processing and pattern recognition technologies.

Research Question:

- What is the importance of sign language as an interpretation language?
- What are the methods/techniques available for Sentence creation utilizing Natural Language processing Engine?

3. METHODOLOGY

3.1.Design:

Hearing and speech impaired users will perform hand gestures based on ISL using body position snippets in the initial stage of the application. A sensor glove is made by the 5DT Company. A little sensor is placed on each finger and thumb. Two additional buttons must be pressed to rotate and tilt the hand. The bend of each sensor is calculated using a value between 0 and 4095. 0 indicates that the sensor is completely stretched, while 4095 indicates that it is entirely bent. These characteristics are compared to the database for each gesture to identify the one-of-a-kind word that corresponds to it. The NLP engine takes those words and gives a sign attribute to each one based on components of speech. These characteristics are then used to construct a meaningful sentence based on the syntax of sentences in English grammar. A grammar must be developed in order to construct sentences.

It involves the following steps:

- A person who is hearing impaired signs.
- As a result of software, signs are translated into text (and video)
- The one who is hearing it reads it (and view it)
- It is important that both hearing and speech impaired talk into a microphone

- Text-to-voice software (and ASL video)

3.2. Instruments:

The movies are captured using a 2D camera. For frame extraction, the largest curvature point is utilized. The photographs with the most distinguishable coefficients are chosen for P2-DHMM training once the DCT coefficients of the images are determined. As a consequence, the data set collected may be used to extract the most important information from the subject's photographs. Using the equation for all the pictures, calculate the distance between each image and chosen training photographs.

$$D_{i,j}^2 = \sum_{n=1}^{N} (d_i(n) - d_j(n))^2$$

Training regarding next image Index = argmax_i (min ($D_{i,j}$)) where 'N' is the length of image vector (No. of rows × No. of columns).

3.3. Data Collection:

Step 1: 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

Step 2: The following formula may be used to divide the video into frames: Framing time = number of frames*FPS as shown in figure 2.

Step 3: For Grayscale conversion, binaries the frames before converting them to black and white as shown in figure 3.

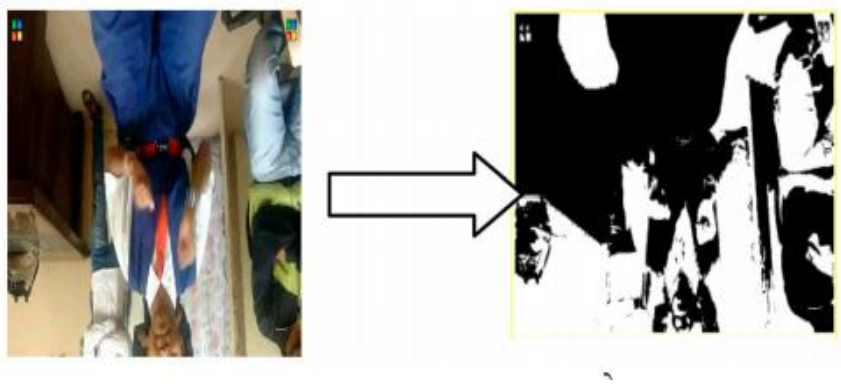


Figure 3: A Single Sample Frame from the Camshift of the Frame into Grayscale i.e. Binary Format

Step 4: Using HSV and CAMSHIFT, segmentation and tracking are conductedas shown in figure 4.

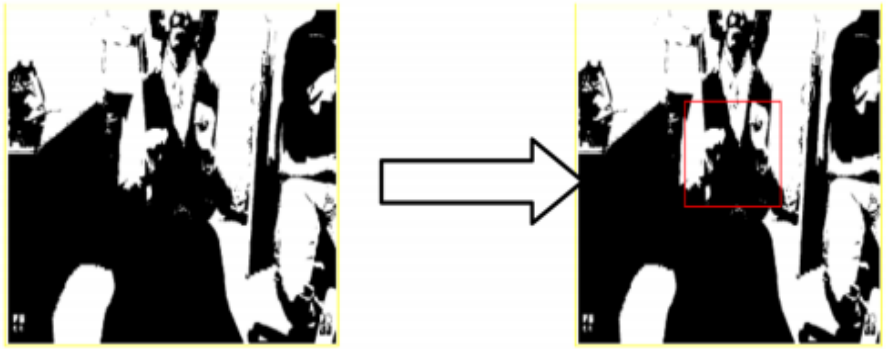


Figure 4: A Single Sample Frame from Grayscale to More Subjected Tracking to the Gesture.

Step 5: 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.

Step 6: With the help of the Haar Classifier, which identifies the model feature using 33 samples of expression, the hand movement is classified and recognized.

Step 7: 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).

3.4. Data Analysis:

Analysis: To provide a sense of the various gesture examples, the samples in the data gathered below were taken from snippets of several movies featuring hand asserted motions. Each of these movies is different in duration. A total of 60 videos were made with various backdrops and lighting conditions. While the number of frames in a video varies with its size, the database sample and video resolution influence the accuracy of a sign language translation system.

TABLE 1: SAMPLE VIDEOS REFERRED TO THE RECORDING AND THE SUCCESS RATE FRAME VISE

Name Of Video File	Sentence In The Video File	Duration of Video	Total Frames	Successfully Recognized Frames From Total Frames
Vid4.avi	Take your glasses off	4 seconds	60	54
Vid3.avi	Think about my idea	5 seconds	56	51
Vid2.avi	Tell me the time	4 seconds	25	19
Vid1.avi	Give me a pen	4 seconds	48	40

TABLE 2: SHORT SYSTEM AND SHORTENED MANUAL GENERATED OUTPUTS SENTENCES/PHRASES.

System Generated Output	Manually Generated Outputs
Take your glasses off	Take glasses off
Think about my idea	Think my idea
Tell me the time	Tell time
Give me a pen	Give pen

Context isn't a factor in any of the grammars we've looked at so far. The initial version of the application's grammar utilized Non-Deterministic Finite Automata. The conversion to Deterministic Finite Automata (DFA) is done in order to enhance the accuracy and results. Superior results were achieved because DFA considers each state as a distinct entity. As a result, the DFA that was generated is clear and simple. It also resulted in the grammar's ambiguity being removed, as well as a substantial increase of the grammar's area of applicability. All kinds of words and tenses are very difficult to comprehend. The software did not accept the tense as an input parameter. As a result, 33.33 percent of the phrases generated from table 2 were correct. This may be illustrated using the following example. The software only works with basic and continuous tenses of speech when it comes to simple and complex sentences. You'll only receive sentences that meet the grammatical rules if you utilize the grammar you've constructed. As a

result, it's conceivable that in certain circumstances, a meaningful sentence will be impossible to come up with.

4. RESULTS AND DISCUSSION

For analytical reasons, the hearing and speech impairment expert analyzed the sentence obtained from the videos. Special school instructors for hearing and speech impairments manually evaluated five films that were translated to sign language using the system's output, yielding five phrases. As a consequence, the output of this system and the manual interpretation are quite similar in terms of outcomes. As a result, it's accurate to within approximately 90% of the time. Hearing and speech challenged individuals may communicate with anybody, anyplace, with this software. This effort also supports speech-to-sign translation. Two of the most essential software components are Video Relay Service and outfit-7 (VRS). It is feasible to create a system that incorporates all of these elements. Nothing has been incorporated in our study as a crucial method of communication among speech disabled individuals except American Sign Language (ASL) (American Sign Language). All letters are signed with just the right hand, with the palm towards the spectator. In terms of manual nature and organization, SE (Sign English) is a good equal to English. People who are deaf or hard of hearing will learn English more rapidly if they are exposed to SE and other parallel signing systems. Sign words and sign markers are the two kinds of gestures used by a sign language interpreter (SE). Each Sign word corresponds to a distinct item in a dictionary of Standard English. As part of our study, we're putting the Sign Word concept into practice, which is useful for translating Sign Language into words, as well as creating Natural Language Processing (NLP) algorithms that can handle longer sentences. With the exception of one, they are signed in the same order as the words in an English sentence. In SE, ASL is the main source of signs. To prevent misunderstanding, these signals are now used in the same sequence and with the same meaning as English terms. Hearing-impaired individuals may also use this software for mobile sign translation (VSR) and UTF-7, which enables them to communicate without calling numbers in everyday activities.

5. CONCLUSION

The main aim of this paper is to illustrate the significance of sign language interpretation and to show how sign language may be translated into words to assist people with hearing and speech impairments in their everyday activities. In this research, just a few words were utilized, and only brief sentences were used to assess the findings. We want to expand our work in the future by creating a large database with the most conceivable words and developing NLP algorithms that can handle longer phrases. As with other languages, converting ISL to English is not a straightforward word-for-word translation. When it comes to recognizing pronouns, prepositions, and other grammatical elements, language learners face a number of difficulties. Prior work on NLP and gesture recognition would be very useful in creating the app. The desire for new software and novel ways to create valuable things will never go away. As a result, an algorithm must be developed in order to build the application. In the future, face emotion detection may be incorporated to the sign recognition software, which would help us punctuate words correctly. This allows us to get a deeper understanding of how someone is experiencing. If we follow specific linguistic principles, we may really contribute to the development of a sign language interpretation system with the help of sign language trainers and hearing and speech challenged

people. Assistive devices for hearing impaired people are required to overcome the communication gap between hearing impaired and normal people.

REFERENCES

1. S. Basar, A. Adnan, N. H. Khan, and S. Haider, "Color Image Segmentation Using K-Means Classification On RGB Histogram," *Recent Adv. Telecommun. Informatics Educ. Technol.*, 2014.
2. M. Mohandes, J. Liu, and M. Deriche, "A survey of image-based Arabic sign language recognition," 2014, doi: 10.1109/SSD.2014.6808906.
3. C. H. Wu, Y. H. Chiu, and C. S. Guo, "Text generation from Taiwanese sign language using a PST-based language model for augmentative communication," *IEEE Trans. Neural Syst. Rehabil. Eng.*, 2004, doi: 10.1109/TNSRE.2003.819930.
4. P. S. Rajam and G. Balakrishnan, "Real time Indian Sign Language Recognition System to aid deaf-dumb people," 2011, doi: 10.1109/ICCT.2011.6157974.
5. M. Zimmermann, J. C. Chappelier, and H. Bunke, "Offline grammar-based recognition of handwritten sentences," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2006, doi: 10.1109/TPAMI.2006.103.
6. A. D. Wilson and A. F. Bobick, "Parametric hidden Markov models for gesture recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, 1999, doi: 10.1109/34.790429.
7. K. Dabre and S. Dholay, "Machine learning model for sign language interpretation using webcam images," 2014, doi: 10.1109/CSCITA.2014.6839279.
8. C. Jung, C. Kim, S. W. Chae, and S. Oh, "Unsupervised segmentation of overlapped nuclei using bayesian classification," *IEEE Trans. Biomed. Eng.*, 2010, doi: 10.1109/TBME.2010.2060486.
9. S. A. Mehdi and Y. N. Khan, "Sign language recognition using sensor gloves," 2002, doi: 10.1109/ICONIP.2002.1201884.
10. M. C. Surabhi, "Natural language processing future," 2013, doi: 10.1109/ICOISS.2013.6678407.