SIGN LANGUAGE INTERPRETATION USING IMAGE AND NATURAL LANGUAGE PROCESSING

Pranali Bhosale*1, Amruta Dhone*2, Sambhaji Jadhav*3, Shreeharsh Sawant*4, A.H.Nadaf*5

*1,2,3,4 ug Student, Computer Science & Engineering, Dr.J.J.Magdum College Of Engineering, Jaysingpur, Maharashtra, India.
*5 asst.Professor, Computer Science & Engineering, Dr.J.J.Magdum College Of Engineering, Jaysingpur, Maharashtra, India.

ABSTRACT

Deaf and dumb people use sign language to communicate. There are various sign recognition techniques that produce output in the form of words or identified signs. The suggested method focuses on sign language interpretation in correct English sentences. In addition to sign identification, several NLP (Natural Language processing) techniques are applied. Input is given as video of sign language followed by framing & segmentation on video. One of the fastest-growing areas of research is sign language recognition. In this field, many innovative techniques have lately been created. Sign language is mostly used by deaf and dumb people to communicate. For persons who are hard of hearing, sign language is the most natural and expressive way to communicate. People who are not deaf, never attempt to learn sign language in order to communicate with deaf people. Deaf people get isolated as a result of this. However, if a computer can be programmed to convert sign language into written format, the gap between normal people and the deaf community can be narrowed. The Haar Cascade classifier is used to identify signs. The continuous words for each sign are sent as input to the POS (Part of Speech) tagging module after sign recognition. It is utilised a Word Net POS tagger with its own Word Net Dictionary. Finally, the sentence is framed using the LALR Parser. In this approach, the suggested sign language interpreter model produces intelligible English sentence.

Keywords: Sign Language, Image Processing, Natural Language Processing.

I. INTRODUCTION

Communication is the process of passing information from one person to another. The majority of the time, people communicate with signs and speech. Normal people utilize natural language to communicate and engage with one another, whereas deaf and dumb persons employ tactile sign language. People with impairments are finding it increasingly difficult to compete in today's world due to fierce competition in every filed. According to report, India has nearly 2.4 million deaf and dumb people, accounting for almost 20% of the world's total deaf and dumb population. For hassle-free interaction between the normal person and deaf and dumb person, there is a need of an interpreter (Person who has the knowledge of sign language, as well as normal language).

Visual sign language and Tactile sign language are the two types of sign language.

1. Visual sign language
   It is for visually impaired i.e. Hearing and speech disabled persons.
2. Tactile sign language
   It is used by those who are deaf and blind.

We are mostly focused on deaf and dumb people's visual sign language. Sign language varies by country and culture. For example, in India, ISL (Indian Sign Language) is used, while in America, ASL (American Sign Language). Sign language is a deaf and dumb communication system made up of a variety of motions generated by hand shapes, body postures, and face expressions. Each gesture has a meaning assigned to it.

Words are made up of hand forms with different orientations. Facial expressions are included in complete visual sign language. Deaf and dumb people can communicate effectively using visual sign language. Though this is true, hearing impaired people must overcome communication barriers in a culture where the majority of people can hear. Visual sign language interaction will be the focus of this study. Natural language is a skill used for understanding human language. It is a part of linguistics and Artificial Intelligence.
NLP is a step in the process of creating a system that can turn text (words) into human language. NLP’s POS tagging method was initially introduced in 1960. It’s crucial technique for language processing. It is the simplest and most stable step in many NLP applications. POS tagging is the first stage in machine translation, information retrieval, and other processes. Second important method in NLP is parsing. Parsing is the method which is followed by the compiler.

II. RELATED WORKS

Different academics identified distinct hand gestures using diverse methodologies, which were then used in diverse fields. Vision-based approaches, data glove-based approaches, and soft computing approaches such as Artificial Neural Networks, Fuzzy logic, Genetic Algorithms, and Canonical Analysis, among others, were used to recognize diverse hand motions. Hand segmentation approaches, feature extraction approaches, and gesture recognition approaches are the three primary types of recognition techniques. The usual technology is introduced in “Application study on face detection technology employs Open CV technology in mobile augmented reality.” Open CV, or open source computer vision library, is a cross-platform library. The Open CV, written in C, provides a very extensive visual processing method, which is integrated with the open source characteristics. To interpret gestures for human-computer interaction, data gloves and vision-based methods are widely utilized. In the data gloves approach, sensors connected to a glove convert finger flexion into electrical signals to determine hand posture. In the vision-based method, the camera is employed to collect image motions. As with the glove-based method, the vision-based method reduces the obstacles. This work introduces American Sign Language standards with the title “Hand talk-a sign language recognition based on accelerometer and semi data.” It’s a part of “deaf culture,” with its own set of puns, inside jokes, and so on. The “Hand gesture detection and voice conversion system for stupid people” proposes bridging the gap between the mute group and the rest of the world. The proposed methodology translates text into speech. The technique helps unintelligent individuals overcome their time constraints and enhance their demeanour. In comparison to the existing system, the proposed configuration is simple and compact, and it can be carried to any location. This technique translates the language into an associate text that ancient people can understand. To assist deaf persons, the language is translated into some form of text that is exhibited on a digital display screen. This approach is useful in the real world for deaf and dumb people who are unable to communicate with normal people.

![Flowchart Image](image-url)

**Fig.1. Flowchart**

III. METHODOLOGY

The recognition of sign language is an essential application of gesture recognition. The following method is used to recognize sign language.
Vision based approach

In a vision-based methodology, image processing techniques are utilized to recognize and track the signer’s hand signs and facial expressions. This method is more convenient for the signer because no additional hardware is required. However, accuracy issues with image processing algorithms exist, and these issues have yet to be addressed.

In terms of vision-based sign language recognition, there are two approaches:

- 3D model based
- Appearance based

Methods based on 3D models make use of 3D information from important body sections. Several key factors, such as palm position, joint angles, and so on, can be calculated using this data. This method employs volumetric, skeleton, or a mixture of the two models. The volumetric approach is more suited to the computer animation and computer vision industries. This method is computationally demanding, and systems for real-time analysis have yet to be developed.

Images are used as inputs in appearance-based systems. These videos/images are directly interpreted by them. They don’t employ a three-dimensional representation of the human body. Using a template database, the parameters are derived directly from the photos or videos. Some of the templates are malleable 2D templates of human body parts, particularly hands. Deformable templates are groups of points on the contour of an item. It serves as an interpolation node for the approximation of an object’s outline.

LDA Algorithm

Linear discriminant analysis (LDA) is the generalisation of Fisher’s linear discriminant (FLD). LDA makes use of projections of training images into a subspace defined by the fisher faces known as fisherspace. Recognition is performed by projecting a new face onto the fisher space. LDA comes with following additional capabilities,

1. LDA can be applied to two or more than two-class classification problems.
2. LDA works better when classes are well separated.

Statistics, pattern identification, and machine learning are the key applications of LDA. It’s used to determine a linear combination of characteristics that distinguishes two or more groups of objects or occurrences. Its combination also used for dimensionality reduction before later classification. LDA is also closely resembles to principal component analysis (PCA) and factor analysis. Both PCA and factor analysis is linear combinations of variables and they describe the data in a better manner. LDA demonstrates how to model the differences in data classes. Factor analysis, on the other hand, constructs feature combinations based on differences rather than similarities, which PCA cannot account for. Discriminant analysis differs from factor analysis in that, it is not an interdependence methodology, requiring a differentiation between independent and dependent variables (also known as criteria variables). For each observation in LDA, the measurements conducted on independent variables are continuous quantities. When dealing with categorical independent variables in LDA, discriminant correspondence analysis is used.

How LDA works

Let’s look at how LDA accomplishes this by projecting characteristics from a higher dimension to a lower dimension space.

1. Computes mean vectors of each class of dependent variable
2. Computers with-in class and between-class scatter matrices
3. Computes eigen values and eigenvector for SW (Scatter matrix within class) and SB (scatter matrix between class).
4. Sorts the eigenvalues in descending order and select the top k.
5. Creates a new matrix containing eigenvectors that map to the k eigen values.
6. Obtains the new features (i.e. linear discriminants) by taking the dot product of the data and the matrix.

Data Acquisition

For each of the 26 signs, ten photos will be acquired in order to attain a high level of accuracy in the sign language recognition system. These photos are stored in a database for training and testing.
The signer adjusts the acquired image at a distance to get the needed image sharpness.

**Pre-Processing**

Image acquisition, segmentation, and morphological filtering procedures are all part of the pre-processing process.

**Image acquisition**

Pre-processing begins with this stage. The process of acquiring an image with a sensor (such as a camera) and transforming it into a controllable entity is known as image acquisition (for example, a digital image file). This is the procedure for detecting a picture. As a result, image acquisition is aided by "illumination." Pre-processing, such as scaling, will also be required. The image will be taken from a database during image acquisition.

**Segmentation**

Segmentation is the technique of dividing an image into small segments in order to extract more accurate picture attributes. The representation and description of an image will be accurate if the segments are appropriately autonomous (two segments of an image should not contain any similar information), but the outcome will not be accurate if rough segmentation is used. Hand segmentation is used here to isolate the object from the backdrop. For segmentation, the Otsu algorithm is utilized. Certain traits are depicted in the segmented hand image.

**Morphological Filtering**

Morphological filtering of a binary image is conducted by considering compound operations like opening and closing as filters. They may act as filters of shape. For example, opening with a disc structuring element smooths corners from the inside, and closing with a disc smooths corners from the outside. Morphological Filtering tools extract the image components, which are useful for form representation and description. The picture attribute is unquestionably the process's result. Gesture recognition is based on the features acquired from the segmentation operation. Morphological filtering techniques are used to remove noise from the photographs, resulting in a smooth contour. The saved database is used to perform the pre-processing step.

Morphological operators-dilate, erode, open, and close, can be applied through image filtering to grow or shrink image regions, as well as to remove or fill-in image region boundary pixels. As areas of foreground pixels shrink in size, holes within those areas become larger.

Erosion and Dilation are the two most basic morphological activities. They can be used in a variety of ways, for example,

- Removing noise
- Isolation of individual elements and joining disparate elements in an image.
- Finding of intensity bumps or holes in an image

**Dilation and Erosion**

Two basic morphological operations are dilation and erosion. Dilation increases the number of pixels on the edges of objects in an image, whereas erosion reduces the number of pixels on the edges of objects. Dilation and Erosion are mostly used for binary images (pixels have values of 1 or 0 only).

- **Dilation**
  - This operations consists of convolving an image \( A \) with some kernel \( B \), which can have any shape or size, usually a square or circle.
  - The kernel \( B \) has a defined anchor point, usually being the center of the kernel.
  - As the kernel \( B \) is scanned over the image, we compute the maximal pixel value overlapped by \( B \) and replace the image pixel in the anchor point position with that maximal value. As you can deduce, this maximizing operation causes bright regions within an image to "grow" (therefore the name dilation).
  - The dilatation operation is: \( \text{dst}(x,y) = \max(x',y') : \text{element}(x',y') \neq 0 \text{src}(x+x',y+y') \)
  - Applying dilation we can get:
- Erosion

- This operation is the sister of dilation. It computes a local minimum over the area of given kernel.
- As the kernel $B$ is scanned over the image, we compute the minimal pixel value overlapped by $B$ and replace the image pixel under the anchor point with that minimal value.
- The erosion operation is: $\text{dst}(x,y) = \min(x',y') : \text{element}(x',y') \neq 0 \Rightarrow \text{src}(x+x',y+y')$

Similarly to the dilation example, we may use the erosion operator on the image (shown above). The bright parts of the image become thinner, while the black zones become larger, as shown in the result.

**Feature Extraction**

Feature extraction, also known as feature selection, is widely utilized in statistical research and machine learning. Choosing a feature to be extracted necessitates a thorough awareness of the domain as well as prior knowledge of the issue at hand. For example, in the field of computer vision imagine that we have a 100X100 pixel image. Then the raw vector intensities become 10000. Often the image corners do not contain much useful information. Dimensionality could be greatly reduced if we accept a modest amount of information loss in exchange for keeping the image pixel in the centre and dropping the pixels in the corners.

**IV. SIGN RECOGNITION**

The LDA methodology for sign recognition is a dimensionality reduction technique that extracts the necessary number of primary components from multi-dimensional input. The LDA algorithm is used to recognise gestures in two steps.

- Training Phase
- Recognition Phase

During the training phase, each gesture is represented as a column vector. After that, the gesture vectors are normalised against the average gesture. The programme then uses a speed-up approach to locate the eigenvectors of the covariance matrix of normalised gestures, reducing the number of multiplications required. The appropriate gesture space projections were produced by multiplying each of the gesture vectors by the eigenvector matrix. A subject gesture is normalised with respect to the average gesture and then projected onto gesture space using the eigenvector matrix during the recognition phase. Finally, the Euclidean distance between this projection and all other projections is calculated. During the training phase, the minimum value of these comparisons is chosen for recognition.

Finally, the identified sign is translated into appropriate text and speech for display on the GUI.
V. EXPERIMENTAL RESULTS

Following figures shows the experimental results. We have taken some of the signs, which are used by deaf and dumb people. To clarify our project motto, we have given four signs and their respected results as shown below:

VI. CONCLUSION

Currently, research efforts have primarily concentrated on recognising static ISL indications from pictures or video sequences captured under controlled conditions. The dimensionality of the sign recognition process will be minimised by employing the LDA method. Noise will be decreased and with excellent accuracy as a result of dimensionality reduction. We attempted to construct this system by combining numerous image processing approaches and fundamental picture features. The recognition of gestures has been accomplished using LDA.
algorithms. Remembering that every God creature has value in society, let us endeavour to incorporate hearing challenged persons in our daily lives and live together.

VII. REFERENCES


