

# Talking Fingers - Sign Language Translator

<sup>1</sup>Akshay Agrawal, <sup>2</sup>Hardik Kundalia, <sup>3</sup>Swati Anakal, <sup>4</sup>Sourabh Kudturkar

<sup>1</sup>Assistant Professor, <sup>2,3,4</sup>Undergraduate Students, Department of Information Technology,

<sup>1,2,3,4</sup>Universal College of Engineering, Kaman, Maharashtra, India

<sup>1</sup>akshay.agarwal@universal.edu.in, <sup>2</sup>hardikkundalia16@gmail.com, <sup>3</sup>anakalswati@gmail.com,

<sup>4</sup>sourabhkudturkar@gmail.com

**ABSTRACT** - Sign languages are visual languages that use hand, facial and body movements as a means of communication. There are over 135 different sign languages all around the world including American Sign Language (ASL), Indian Sign Language (ISL) and British Sign Language (BSL). Sign language is commonly used as the main form of communication for people who are Deaf or hard of hearing, but sign languages also have a lot to offer for everyone. In our proposed system, we are creating a Web Application which contains two modules: The first module will accept the Information in Natural Language (Input Text) and it will show the corresponding Information in Sign Language Images (GIF Format). The second module will accept the Information in Sign Language (Input Hand Gesture of any ASL Letter) and it will detect the Letter and display it as the Output (Text). The system is built to bridge the communication gap between deaf-mute people and regular people as those who don't know the American Sign Language can either use it to learn the Sign Language or to communicate with someone who knows the Sign Language. This approach will help users in quick communication without having to wait for any human interpreter to translate the Sign Language. The application is developed using Django and Flask frameworks and it includes NLP and Neural Network. We are focusing on improving the Living standards of the hearing impaired people as it can be very difficult to perform everyday tasks especially when people around them don't know Sign Language. This application can also be used as a teaching tool for relatives and friends of deaf people as well as people interested in learning the sign language.

**INDEX TERMS** - *Sign Language, American Sign Language, Text to Sign, Sign to Text, Gesture Recognition, Natural Language Processing, Neural Network.*

## I. INTRODUCTION

Sign languages are visual languages that use hand, facial and body movements as a means of communication. It is not a universal language which means every country has its own sign language. There are over 135 different sign languages all around the world including American Sign Language (ASL), Indian Sign Language (ISL) and British Sign Language (BSL). There are also signed representations of oral languages such as Signed Exact English (SEE) and mixes such as Pidgin Signed English (PSE). Sign language is commonly used as the main form of communication for people who are Deaf or hard of hearing, but sign languages also have a lot to offer for everyone. WHO estimates that there are 466 million persons in the world with disabling hearing loss (6.1% of the world's population) out of which 432 million (93%) of these are adults (242 million males, 190 million females) and 34 million (7%) of these are children [9]. Projections show that the number could rise to 630 million by 2030 and may be over 900 million in 2050. American Sign Language (ASL) is a complete, natural language that has the same linguistic properties as spoken languages, with grammar that differs from English. ASL is expressed by movements of the hands and face. ASL is a language

completely separate and distinct from English. It contains all the fundamental features of language, with its own rules for pronunciation, word formation, and word order. While every language has ways of signalling different functions, such as asking a question rather than making a statement, languages differ in how this is done. For example, English speakers may ask a question by raising the pitch of their voices and by adjusting word order; ASL users ask a question by raising their eyebrows, widening their eyes, and tilting their bodies forward. American Sign Language (ASL) plays an important role in facilitating communication within the Deaf community and enabling them to communicate with others. However, did you know that sign language is growing in popularity as a second language around the world? ASL is not a language limited to the Deaf. It is growing in popularity with an estimated 2 million speakers in North America alone. It is the 3rd most commonly studied language in the United States after English and Spanish.

## II. RELATED WORKS

T. Vichyaloetsiri and P. Wuttidittachotti [1] in this paper, translation of text into sign language animated GIF is proposed using a new Framework. This framework is developed for Thai Sign Language. There are 42,121

words, 30 sign language images in the dictionary. The framework makes the use of web service consisting of REST and JSON. It uses longest word division technique to extract the words from the sentence. It brings together all the images into a GIF format with a frame rate of 20-30 frames per second. The framework was tested on Postman Chrome App, PHP Website and created Mobile Application and same results were obtained on each application. There were 10 translation errors while 40 were correct resulting in 80% accuracy of the framework. Security of data in transit without encryption during transmission was the main concern and the integrity of the animation depends on the source image added to the dictionary.

M. Papadogiorgaki, N. Grammalidis, D. Tzovaras and M. G. Strintzis [2] in this paper, translation of text into sign language is proposed. This application is developed for American Sign Language. The application takes text as input and provides a output in the form of 3D animated VRML sequences which can be viewed on any VRML-compliant browser. The VRML animation sequences are generated from Sign Language Notation based on MPEG-4 Face and Body Animation. The application was tested and the problem of synthesis of static and dynamic gestures was identified. A major problem identified was that when the sign box contains contact symbols, the touch between the hands or the hand and the face is difficult to achieve.

M. Grif and Y. Manueva [3] in this paper, translation of text into sign language is proposed. The system is developed for Russian Sign Language. The system uses Morphological analysis of words, Syntactic analysis and Semantic analysis of words. The original sentence is divided into two parts, each of which represents a smaller number of cases for simpler comparisons. The main disadvantage of the system is lack of consideration of the semantic component of both sound language and sign language. To increase the quality of the sign language interpreter, software unit is developed.

N. Aouiti [4] in this paper, translation of text into sign language is proposed. The system is developed for Arabic Sign Language. The system uses Morpho-syntactic analysis, Semantic analysis and Pragmatic analysis. Arabic language is composed of 28 letters and it is written from right to left. Some letters may change shape depending on their position in the word. The system uses Stanford Parser for linguistic treatment which is available for free. The construction of dictionaries of signs is an expensive operation and needs a lot of time.

Amit Kumar Shinde and Ramesh Kagalkar [5] in this paper, translation of sign to text as well as text to sign, both methods are proposed. This system is developed for Marathi Sign Language. The system contains 43 Marathi Sign Language alphabets and around 500 words and the database contains 1000 different gesture images. The

system performs preprocessing of input which includes gray scale convergence, edge detection, generation of array of image, compare or matching with database. For edge detection, Canny's edge detection algorithm is used. The system is developed in Java and uses NetBeans as platform. It is a low time consumption approach. The time required to translate sign language to text is more than the time required to translate text to sign, because of the time required to process image takes more time than text.

Krishna Modi and Amrita More [6] in this paper, translation of sign language into text is proposed. This system is developed for American Sign Language. The system captures the video frame-by-frame and then it is processed to extract the appropriate image which is further processed using BLOB analysis and sent to statistical database. The processing of the image includes smoothening of the Image and converting it into grayscale. The captured image is compared with the images saved in the database and the matched image is used to determine the performed alphabet sign. The constraint of the system is that the background of every captured image should be of a distinctively different color like black which can be easily differentiated. The finger-spelling should not contain any type of movement so the system does not support the letters J and Z. The image should only consist of the single hand without any accessories or any other body part. The system uses the AForge.NET library and it is implemented in Visual Studio 2010 using C#. The system was tested and gave 96% accuracy in the results. The system takes at most 50 milliseconds to give the output which is the least possible processing time. It can also be used to from words as well as sentences.

NB, Mahesh Kumar [7] in this paper, translation of sign language into text is proposed. This system is developed for Indian Sign Language. The system contains 10 images of 26 signs each. The system has 2 approaches for gesture recognition. Glove based approach makes use of a sensor glove. Vision based approach makes use of Image Processing algorithms to detect and track hand signs and facial expressions. LDA Algorithm is used for gesture recognition. Pre-processing includes image acquisition, segmentation and morphological filtering methods. Dilation and Erosion are used for binary images. The effect of erosion is to shrink objects by removing the pixels from the boundary of the object. The effect of dilation is to increase the size of an object by adding pixels. The reduction of data dimensionality is called as Feature extraction technique. Sign Recognition using LDA algorithm involves two phases i.e. Training phase and Recognition phase. Each gesture is represented as a column vector in the training phase. A subject gesture is normalized with respect to the average gesture in the recognition phase. By using LDA algorithm, the dimensionality will be reduced due to which the noise will be reduced and high accuracy will be obtained.

Anand, M. Suresh, A. Kumaresan, and N. Mohan Kumar [8] in this paper, translation of sign language into voice as well as voice to sign language, both methods are proposed. This system is developed for Indian Sign Language. The continuous images of hand movements are taken as the input from the device which is pre-processed, classified and filtered and then image feature vector is created. It is compared with the database to find the meaning of the sign. The recognized sign's meaning is sent to the voice generator and the voice is transmitted to the receiver. Voice input is converted to text by using trained voice database. The text is compared with the database for finding the meaning and sign. This sign is now displayed to the receiver. The challenge in implementing this system on mobile phone is to implement the image processing methods.

### III. AMERICAN SIGN LANGUAGE

American Sign Language (ASL) is a complete, natural language that has the same linguistic properties as spoken languages, with grammar that differs from English. ASL is expressed by movements of the hands and face. It is the primary language of many North Americans who are deaf and hard of hearing, and is used by many hearing people as well [10]. Fingerspelling is part of ASL and is used to spell out English words. In the fingerspelled alphabet, each letter corresponds to a distinct handshape. Fingerspelling is often used for proper names or to indicate the English word for something.



Fig. 1. American Sign Language images

Figure 1 shows the sign language images for corresponding ASL alphabets. Proposed system is designed to recognize the 26 ASL letters. During the recognition ASL is translated into corresponding English text and similarly for vice versa.

ASL has a productive system of classifiers, which are used to classify objects and their movement in space. For example, a rabbit running downhill would use a classifier consisting of a bent V classifier handshape with a

downhill-directed path; if the rabbit is hopping, the path is executed with a bouncy manner. In general, classifiers are composed of a "classifier handshape" bound to a "movement root". The classifier handshape represents the object as a whole, incorporating such attributes as surface, depth, and shape, and is usually very iconic. The movement root consists of a path, a direction and a manner.

### Fingerspelling

ASL possesses a set of 26 signs known as the American manual alphabet, which can be used to spell out words from the English language. Such signs make use of the 19 handshapes of ASL. For example, the signs for 'p' and 'k' use the same handshape but different orientations. A common misconception is that ASL consists only of fingerspelling; although such a method (Rochester Method) has been used, it is not ASL.

Fingerspelling is a form of borrowing, a linguistic process wherein words from one language are incorporated into another. In ASL, fingerspelling is used for proper nouns and for technical terms with no native ASL equivalent. There are also some other loan words which are fingerspelled, either very short English words or abbreviations of longer English words, e.g. O-N from English 'on', and A-P-T from English 'apartment'. Fingerspelling may also be used to emphasize a word that would normally be signed otherwise.

### IV. METHODOLOGY

The system is useful for two way communication i.e. Text to sign and Sign to text. The system is implemented using python and it is divided into two parts.

#### Block Diagram

Figure 2 shows the block diagram of the first part of the proposed system where the input is in the form of Text and the result is the animated GIF of the equivalent input text.

Figure 3 shows the block diagram of the second part of the proposed system where the input is in the form of Gesture and the result is the English letter of the detected gesture.

#### Text to Sign Conversion

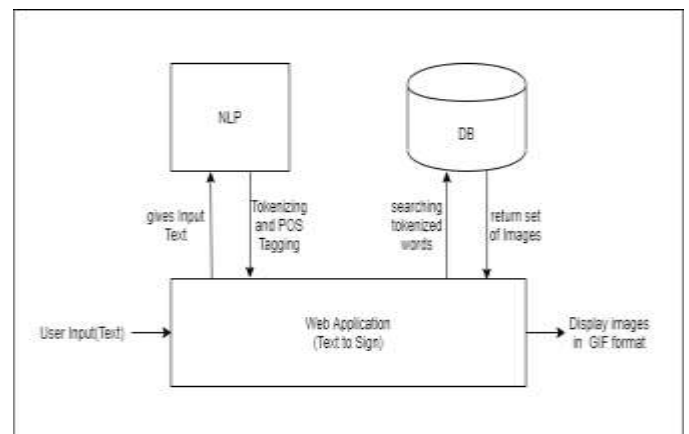


Fig. 2. Block diagram (Text to Sign conversion)

## Input Text

The input text entered by the user is stored in the system. It can be in the form of an alphabet, a word or a sentence. The input text is entered into the system through keyboard.

## Natural Language Processing

Natural language processing (NLP) is a subfield of linguistics, computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data. The Natural Language Toolkit, or more commonly NLTK, is a suite of libraries and programs for symbolic and statistical natural language processing (NLP) for English written in the Python programming language. This toolkit is one of the most powerful NLP libraries which contains packages to make machines understand human language and reply to it with an appropriate response. Tokenization, Stemming, Lemmatization, Punctuation, Character count, word count are some of these packages. The input text might contain some punctuation marks. To remove the punctuation marks from the input text, we are going to use RegexpTokenizer. Tokenizing and removing all punctuation marks from a sentence removes all punctuation marks from each word. The result of this step is a list containing individual words without any punctuation marks.

## Database

The database contains images of the alphabets, numbers, and some words. After obtaining the list of tokenized words, we can match each individual word with the existing words in the database. If the word is matched with the database, it is kept as it is. But if the word does not match with the database, then the word is split into individual letters. The result of this step is a list which contains either words or alphabets or combination of both.

## GIF

GIFs can be used for small animations and low-resolution video clips. Imageio is a Python library that provides an easy interface to read and write a wide range of image data, including animated images, video, volumetric data, and scientific formats. The previous list is given as the input to Imageio which loads all the equivalent images of words or alphabets in the same sequence. Simultaneously, it also appends all the images together with a set duration time, creating a GIF file.

## Output

The final result is the GIF File which is displayed to the user as the system output.

## Sign to Text Conversion

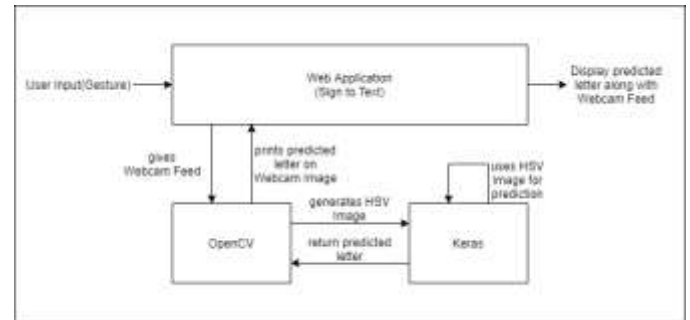


Fig. 3. Block diagram (Sign to Text conversion)

## Input Gesture

The input gesture entered by the user is stored in the system. This input gesture has to be an ASL alphabet which is entered into the system through webcam. The webcam can be internal as well as external. The input taken is in the form of a live webcam feed. This webcam feed is processed frame-by-frame by OpenCV.

## OpenCV

OpenCV or Open source computer vision is a library of programming functions mainly aimed at real-time computer vision. OpenCV-Python is a library of Python bindings designed to solve computer vision problems. OpenCV-Python makes use of Numpy, which is a highly optimized library for numerical operations with a MATLAB-style syntax. All the OpenCV array structures are converted to and from Numpy arrays. In color image processing, there are various models one of which is the hue, saturation, value (HSV) model. Using this model, an object with a certain color can be detected and to reduce the influence of light intensity from the outside. Hue is a measure of the wavelength found in the dominant color received by the sight while the Saturation is the size of the amount of white light mixed in hue. Value works in conjunction with saturation and describes the brightness or intensity of the color. We use OpenCV to capture the Webcam feed and store it in our System. The webcam feed is stored in frames. There is a set period of gap between consecutive frames. We use in-built OpenCV functions to perform a few tasks. We use rectangle function to draw a rectangle on the frame and crop the contents inside the rectangular frame. This rectangular box is where the user will show the ASL gesture. This new rectangular image is now converted into a HSV Image using cvtColor function. The values of HSV are adjusted in such a way that the resultant HSV image looks like a Black and White image, where the Background is pure Black and the Hand Gesture is pure White. This HSV image is saved and processed for each Webcam frame. We also use putText function to print the predicted letter on the frame along with the rectangle.

## Keras

Keras is an open-source neural-network library written in Python. Keras contains numerous implementations of commonly used neural-network building blocks such as layers, objectives, activation functions, optimizers, and a host of tools to make working with image and text data easier to simplify the coding necessary for writing deep neural network code. In addition to standard neural networks, Keras has support for convolutional and recurrent neural networks. In deep learning, a convolutional neural network is a class of deep neural networks, most commonly applied to analyzing visual imagery. We use in-built Keras functions to perform a few tasks. We can use classifier.add function with various parameters to create a new Convolution Neural Network if we want to make the model for a different set of images or to make changes to the existing model. We use load model function to load the model into the system before the webcam feed starts. Once the model is loaded, we get the webcam feed from OpenCV, and the HSV image is saved. We use load img function to load the HSV image into the system. Once the HSV image is loaded, we use classifier.predict function to predict the Letter, wherein the HSV image and the model are taken as input. The result of this step is the letter which is predicted by the Convolution Neural Network which is then printed by OpenCV on the webcam feed.

## Output

The final output of the system is the live webcam feed with the rectangle printed on it. When the user places their hand inside the rectangle, the predicted letter is also printed on the webcam feed.

## V. RESULT AND ANALYSIS

The proposed system is implemented using Python and it uses Django framework as well Flask framework. In the result and analysis, we will look at the working of system through snapshots.



Fig. 4. Input of Text to Sign

In Figure 4, the input text which has to be converted into sign language has to be entered. The input text can either be a letter or a set of letters, a word or a set of words or the

combination of both i.e. a sentence and the output is displayed as given below.



Fig. 5. Output of Text to Sign in Gif Format

In Figure 5, the output displays all the images of words and letters sequentially in a GIF format with a set duration period after which each images is displayed. Earlier, the output used to display all the images at the same time in a sequence but it would take up a lot of space on the screen so now the GIF takes only small amount of space on the screen but it displays all the images one after the another.



Fig. 6. Output of Sign to Text

In Figure 6, the detection and the translation of a letter from sign to text is shown. The detected letter is shown on the live webcam feed itself along with the detection rectangular box. This detection and translation is possible because of conversion of the original image into a HSV image as shown below.

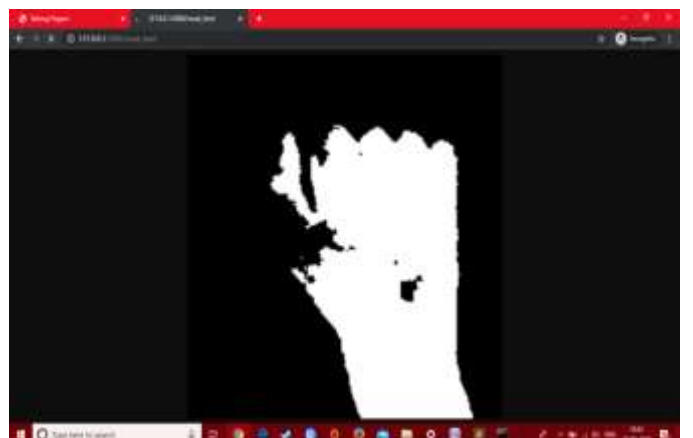


Fig. 7. Output of HSV Image

In Figure 7, the original image is converted into a HSV image so that it can be compared with the existing dataset to predict the letter which the user is trying to show.

## VI. CONCLUSION AND FUTURE SCOPE

The proposed system applies a simple and fast method which works effectively to recognize the hand gestures as well as text. The dataset contains the various hand gestures of different ASL letters that are captured in HSV image format. The system is trained on predefined model which includes 26 ASL letters. In the system, the HSV image is used to recognize the input sign gesture. The process of sign to text conversion consumes very less time, so the recognition is achieved easily in very less time, while the process of text to sign depends on the size of the input i.e. the number of characters or words entered. For e.g., for sign to text conversion it hardly takes half a second for the gesture to be detected and updated on the screen. On the other hand, for text to sign conversion, it takes around 2 seconds if 20 characters are given in the input while it takes around 5 seconds if 100 characters are given in the input. The system is capable of handling the different input signs and translates them into text. The system gives accurate results for most of these 26 letters. The system can take more time to process the text and to generate the animated GIF if large input is given.

In the future scope of this project, we can include voice based function. The user will have to give voice input and the system will convert it into sign language or the user can give gesture input and the system will give a voice output of letters. We can extend the functionality of the system by adding more words and their sign images. We can also add numbers and words for detection. Also, we can add different sign languages like ISL (Indian Sign Language), BSL (British Sign Language), etc.

## VII. ACKNOWLEDGEMENTS

I would like to take the opportunity to express my heartfelt gratitude to the people whose help and co-ordination has made this project a success. I thank our HOD Prof. Yogita Mane, Project guide Prof. Akshay Agarwal and co-guide Prof. Aditi Malkar for knowledge, guidance and co-operation in the process of making this project.

I owe project success to my guides and convey my thanks to them. I would like to express my heartfelt to all the teachers and staff members of Information Technology department for their full support. I would like to thank my principal for conducive environment in the institution.

I am also grateful to the library staff of UCOE for the numerous books, magazines made available for handy reference and use of internet facility.

Lastly, I am also indebted to all those who have indirectly contributed in making this project successfully.

## REFERENCES

- [1] T. Vichaloetsiri and P. Wuttidittachotti, "Web Service framework to translate text into sign language," 2017 International Conference on Computer, Information and Telecommunication Systems (CITS), Dalian, 2017, pp. 180-184.
- [2] M. Papadogiorgaki, N. Grammalidis, D. Tzovaras and M. G. Strintzis, "Text-to-sign language synthesis tool," 2005 13th European Signal Processing Conference, Antalya, 2005, pp. 1-4.
- [3] M. Grif and Y. Manueva, "Semantic analyses of text to translate to Russian sign language," 2016 11th International Forum on Strategic Technology (IFOST), Novosibirsk, 2016, pp. 286-289.
- [4] N. Aouiti, "Towards an automatic translation from Arabic text to sign language," Fourth International Conference on Information and Communication Technology and Accessibility (ICTA), Hammamet, 2013, pp. 1-4.
- [5] Amit Kumar Shinde and Ramesh Kagalkar, "Sign Language to Text and Vice Versa Recognition using Computer Vision in Marathi." IJCA Proceedings on National Conference on Advances in Computing NCAC 2015(1):23-28, December 2015.
- [6] Krishna Modi and Amrita More, "Translation of Sign Language Finger-Spelling to Text using Image Processing." International Journal of Computer Applications 77(11):32-37, September 2013.
- [7] Mahesh Kumar N B, "Conversion of Sign Language into Text." International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 9 (2018) pp. 7154-7161.
- [8] Anand, M. Suresh, A. Kumaresan, and N. Mohan Kumar. "An Integrated Two Way ISL (Indian Sign Language) Translation System—A New Approach." International Journal of Advanced Research in Computer Science 4.1 (2013).
- [9] <https://www.who.int/deafness/estimates/en/>
- [10] <https://www.nidcd.nih.gov/health/american-sign-language>