

Audio and Vision based Indian Sign Language Translator

Ms. Jyotsna B N¹, Ms. Sudreasha Das², Ms. Sakshi Singh³, Mr. Amrit Pratyay⁴ and Prof. Kanaiya V Kanzaria⁵

¹⁻⁴Student, School of Computer Science and Engineering, REVA University, Bengaluru, India
Email: jyotsnabn23@gmail.com

⁵Asst. Prof., School of Computer Science and Engineering, REVA University, Bengaluru, India
Email: kanaiya.vk@reva.edu.in

Abstract—There is a need for a real-time two-way communication system to help deaf people communicate with non-disabled individuals by converting Indian sign language to spoken English speech and vice versa. Although various technological advancements have been made over the years to help the deaf community communicate with others, they have failed due to their complexity and limited sign language datasets. The proposed system is a portable hardware device consisting of input and output components, including a camera, microphone, speaker, and LCD screen, connected to a Raspberry Pi. The system uses machine learning and python libraries to convert the deaf individual's gestures into speech for non-disabled individuals and vice versa. The model is trained with a custom dataset to be compatible with Indian sign language.

Index Terms— Linguistics, Indian sign language, Handheld device, CNN, Raspberry Pi.

I. INTRODUCTION

Sign language is a communication system using visual gestures and sign language used by deaf and mute people. In general, it is okay for two hearing impaired people to communicate using a common sign language. The real problem arises when a non-deaf person wants to communicate with a hearing-impaired person. In such scenarios, sign language is required. Sign language is the mother tongue of deaf people. They can communicate well using sign language, but have a lot of trouble when trying to communicate with listeners, especially those who are not proficient in sign language. However, there are communication barriers between deaf and hard of hearing people because these languages are not generally known outside of these communities.

Sign language, which includes facial expressions and gestures known as non-manual sign language, plays an important role in understanding the correct meaning of sign language. Manual signatures are manual patterns and contain original information, while non-manual signatures are important for legibility. Today, between 138 and 300 different sign languages are used worldwide. Each country has its own sign language, such as American Sign Language (ASL), Argentine Sign Language (LSA), British Sign Language (BSL), and Indian Sign Language (ISL). Speech impaired people prefer sign language which is used in their region.

The deaf community in India uses Indian sign language (ISL). It consists of both word level gestures and fingerspelling. Fingerspelling is used to form words with letter by letter coding. The fingerspelling in Indian sign language consists of both static as well as dynamic gestures which are formed by the two hands with arbitrarily complicated shapes. To bridge the communication gap between speech impaired people and normal people, a

two way communication system can be built that converts sign language to speech and speech to sign language.

II. LITERATURE STUDIES

Sign language recognition systems are broadly classified into two categories: hardware based systems and vision based systems. Hardware based systems require the user to wear some devices to extract features describing the hand sign. Cyber glove is a device which is used for extracting the features such as orientation, movements and colour, of the hands. It is widely used for sign language recognition. Vision based systems use digital image processing techniques to extract features and recognize sign. The following papers discuss these ideas.

Deepali G et al. [1] proposed a system using SVM Classifier and PCA. They created a dataset containing 192 images, removed low-frequency background noise during pre-processing, normalized the intensity of individual particle images, and converted the images to grayscale. Then the image was subjected to skin detection and image thresholding to extract the desired part from the image. Then we used PCA for feature extraction. It performed better than LDA and also reduced the dimensionality. SVM is a support vector machine used for supervised learning models, with associated algorithms that use extracted features to analyze the data used for classification and analysis.

S Hore et al. [2] proposed a method for sign language recognition including pre-processing, data cleaning, and data normalization. The important features for classification of the dataset were carefully extracted from the dataset using effective features. After the pre-processing step, data cleaning is required to remove missing values and noise problems. Different techniques such as artificial neural networks, genetic algorithms and particle swarm optimization were used to train the model. The proposed work is tested on a dataset containing 22 ISL gestures and 10 images for each gesture.

Johnny et al. [3] proposed a method using sensor gloves consisting of 5 sensors on each finger to measure the tension between the knuckle and the first joint of the finger and 2 sensors to obtain the tilt and rotation of the hand. The neural network created is used to map the data to its corresponding character. The system can recognize the 24 alphabets in the English language and two punctuation symbols. Real-time translation of the hand gestures into its corresponding alphabet or digit is done using the system. Various machine learning techniques like k-Nearest Neighbours, Decision Tree Classifier, and Neural Network were used to perform the classification.

J. Rekha et.al. [4] used a feature descriptor which is a combination of shape, texture and local movements of hand features for the recognition of Indian sign language alphabet. The shape, texture and finger features of each hand are extracted using Principle Curvature Based Region (PCBR) detector. Wavelet Packet Decomposition (WPD-2) and complexity defects algorithms respectively for hand posture recognition process. To classify each hand posture, multi class non-linear support vector machine (SVM) was used.

V. Adithya et al. [5] proposed a neural network-based method for automatically recognizing the fingerspelling in Indian sign language. The signs are identified by the features extracted from the hand shapes. They used skin colour-based segmentation for extracting the hand region from the image. A new shape feature based on the distance transform of the image is proposed in this work. The features are extracted from the sign image by distance transformation and are used to train a feed forward neural network that recognizes the sign.

D. Deora et al. [6] proposed a framework for Indian Sign Language Recognition system. Segmentation is colour based i.e. colour the gloves (red and blue) are segmented. Recognition is done using PCA for live data frames where every 20th frame is given as input to be recognized. The result of recognition is 94% i.e. the system can recognise 94% of the signs made correctly. The signs in which both the hands overlap each other and the signs which involves motion in them poses a problem in recognition. Fingertip algorithm is used to make database on the basis of number of finger tips. This can be combined with PCA in such a way that on the basis of number of finger tips in an image, PCA compares the input image with the new database formed with the help of finger tips.

N. R. Albelwi et al. [7] proposed a method that uses predefined Haar classifiers to detect the hand's position and a set of processes to prepare the image for the feature 500 extraction task. After normalizing the bounded image, skin detection is applied to detect skin colour in HSV colour space. Transforming images into frequency domain using Fourier Transform enabled them to form the entire signs database into a single vector based on FDs to ease the matching process. This scheme for feature extraction allowed for the use of a simple classification technique, KNN. Classification using the proposed methodology achieved up to 90.55 % recognition accuracy at real time.

III. PROBLEM FORMULATION

The audio and vision based Indian sign language translation device must be capable of converting Indian sign language to speech and vice versa. Fig. 1 shows the architectural diagram of the device and a step- by-step procedure must be followed to achieve the end result of the problem.

Since it is a two-way communication device there will be two ways to interact with the device depending on the user.

A. Audibly impaired user

If the user is an audibly impaired person then the procedure starts with the camera capturing the hand gestures of the user. This camera should be attached to the CSI camera port of Raspberry Pi . These images have to be processed by the software integrated with the device through the camera.

The software should consist of a perfectly trained model that will be able to process the images and signify what character they represent. The result is then given out by the speaker in the form of speech. This speaker is connected to an USB port. The recipient who is an abled person will be able to understand the speech.

B. Abled user

If the user is a person with no disability then they must speak through the microphone connected to the USB port of Raspberry Pi. The microphone is connected to the software and the speech is processed by it such that it can be converted to characters that can be displayed. The characters are then displayed on the LCD screen connected to the HDMI port. The audibly impaired person will be able to view the output and understand what the user is saying.

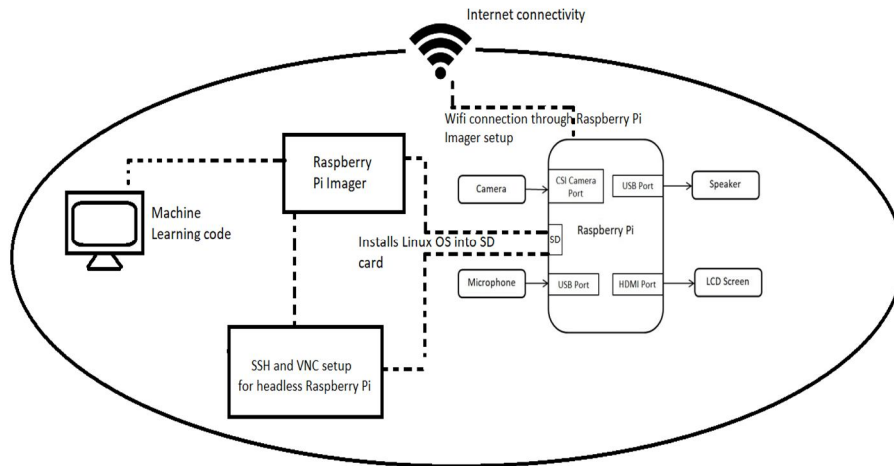


Figure 1: Architectural Diagram

IV. SOLUTION METHODOLOGIES

In order to create an Indian sign language two-way communication device, the software and hardware has to be integrated as shown in Figure 2. The first few steps of collecting the dataset, pre-processing and feature extraction have to be done in order to have a clean dataset that would be easy to give as input to the model for training and testing purpose.

A. Building a Dataset

The lack of availability of public datasets on the internet prompted us to create our own dataset which is the backbone of our project. It would also be made available to the public for future research purposes. The dataset consists of images of the 26 English alphabets along with numbers ranging from 0-9. Approximately 10,000 images are included in the dataset.

B. Pre-processing

Pre-processing allows the images to have less complexity in them and easy for the model to understand. Coloured images are more complex than grey scaled images as they have different channels of colours. Hence pre-processing techniques such as grey scaling and skin detection are used.

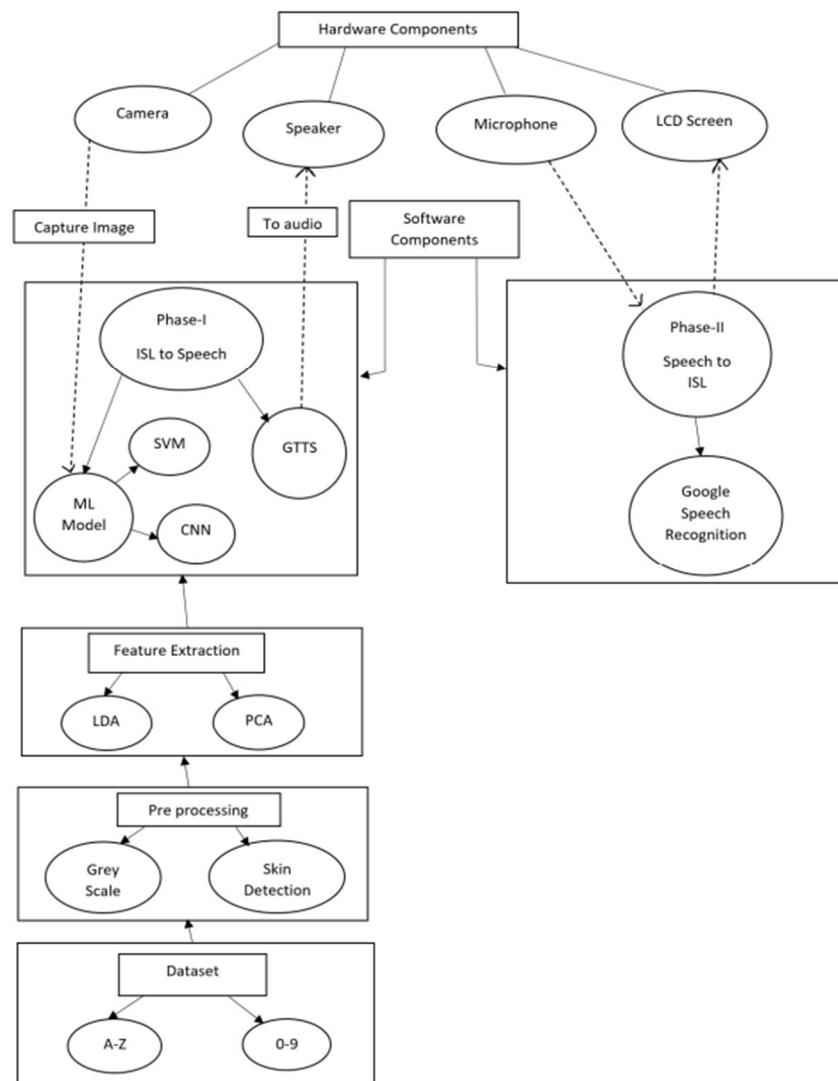


Figure 2: Solution Flowchart

C. Feature extraction

After pre-processing, extracting the clean dataset becomes an easier process. Images which do not have much clarity and may cause confusion while training the model are removed in this process with the help of Linear discriminant analysis (LDA) or Principal Component Analysis (PCA).

The software processing which is done within the device requires the following step by step procedures:

Our proposed system consists of two main phases:

Phase 1: Indian sign language to speech

Image capturing

In order to convert Indian Sign Language to speech, the first step is to capture the images of the symbols shown by the impaired person into the device with the help of a good quality camera.

Data pre-processing

After the images have been perfectly captured, they have to be pre- processed so that it can be easily understood by the algorithm and removes noise. Image processing involves grayscale or RGB conversion and skin detection.

Feature Extraction

The next step involves feature extraction where we will be using LDA (Linear Discriminant Analysis) and PCA (Principal Component Analysis).

Right Machine Learning model selection

After extracting relevant data, the machine learning model is built. The proposed machine learning models as of now are Convolutional neural network (CNN), Support vector machine (SVM), K-nearest neighbour (KNN) and Decision tree. These models would be trained on a custom dataset we have built.

Text to audio conversion

The machine learning model provides individual characters as output which is then converted to audio using Google text to speech library.

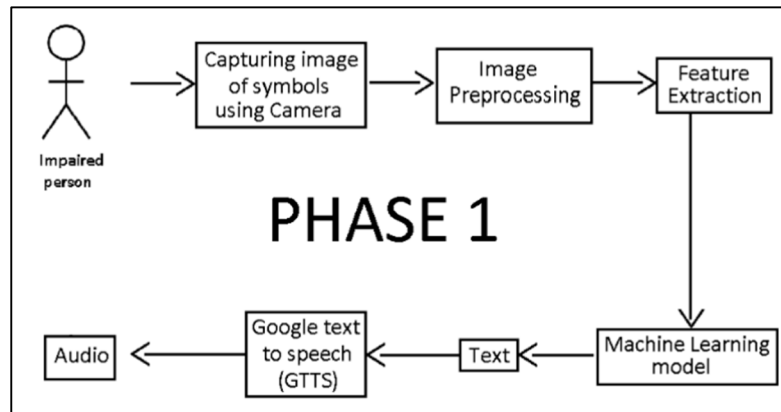


Figure 3: Phase 1- Indian sign language to speech

Phase 2: Speech to Indian sign language

Audio input

In order to convert speech to Indian Sign Language the audio of an individual is taken as an input into the device with the help of a good quality microphone.

Audio to text conversion

The obtained audio is then converted into text with the help of Google speech recognition libraries.

Characters displayed on LCD

The individual characters of the word are then displayed on the LCD screen of the device.

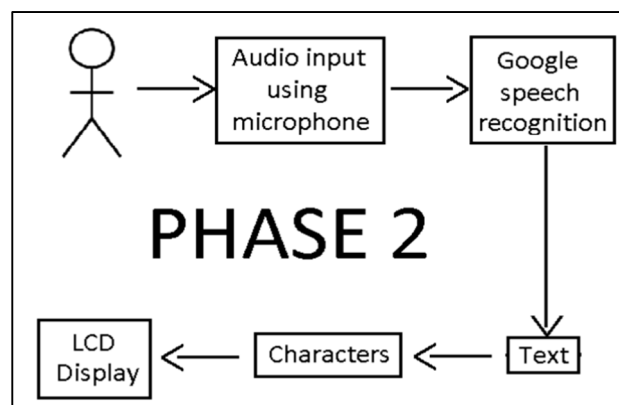


Figure 4: Phase 2- Speech to Indian sign language

V. RESULTS AND SENSITIVITY ANALYSIS

The hardware device that has been proposed in this paper takes in input in two different ways and also gives out the result based on the type of input given. The way of the input provided depends on the user who is interacting with the device.

A user who is audibly impaired and wants to communicate with a disability free user, gives the input to the device through the camera which captures the gestures made by the audibly impaired users. These inputs range

from numbers 0-9 and alphabets A-Z in Indian sign language symbols. The result is given in the form of audio to the user which is also in the same range as the input. The alphabet or number shown by the audibly impaired user translates to its respective English alphabet format and is pronounced in the English language when given as output.

A user who is disability free and wants to communicate with an audibly impaired user, gives the input to the device through the microphone as audio. These inputs range from numbers 0-9 and alphabets A-Z pronounced in the English language. The result is given in the form of English alphabets displayed on the LCD screen which are in the same range as the input. The alphabet or number spoken by the user is translated to its respective English alphabet when given as output.

The sensitivity in our project is mainly due to the occurrence of two different outputs through the same input. Quite a few symbols of the Indian sign language are the same for more than one character. For example, the letter V and number 2 have the same symbol in the Indian sign language which may lead to errors in producing the right output.

Some Indian sign language alphabet symbols are very similar to one another but they mean completely different alphabets. For example, alphabets D and P have similar gestures and may cause the model to interpret them wrongly by interchanging them.

The analysis to this sensitivity can be done based on the entire word or sentence passed to the device. Based on the word the model can interpret the symbol to be a certain alphabet. Words containing alphabets which could confuse the model must be passed to the device more frequently to make changes to the model for more accuracy.

VI. CONCLUSION

The proposed system, audio and vision based Indian sign language translation device, extends some of the capabilities of the works related to Indian sign language recognition by preparing a dataset, integrating the machine learning model with the Raspberry Pi and helping the users to interact with the device with the help of input and output devices connected to it.

The proposed system uses a custom dataset containing approximately 200 images each of the English alphabets and numbers that can be used for future purposes in various other researches and projects related to Indian sign language.

REFERENCES

- [1] Deepali Mali, Nitin Limkar and Satish Mali "Indian Sign Language Recognition using SVM Classifier". Proceedings of International Conference on Communication and Information Processing (ICCIP), May 2019.
- [2] Hore, S., Chatterjee, S., Santhi, V., Dey, N., Ashour, A. S., Balas, V. E., & Shi, F. "Indian sign language recognition using optimized neural networks", Information technology and intelligent transportation systems, Springer, Cham, Dec 2015.
- [3] S. Johnny and S. J. Nirmala, "Sign Language Translator Using Machine Learning", SN Computer Science, vol. 3, no. 1, Jan 2022.
- [4] J. Rekha, J. Bhattacharya and S. Majumder, "Shape texture and local movement hand gesture features for Indian Sign Language recognition", 3rd International Conference on Trendz in Information Sciences & Computing (TISC2011), Dec. 2011.
- [5] V. Adithya, P. R. Vinod and U. Gopalakrishnan, "Artificial neural network based method for Indian sign language recognition," 2013 IEEE Conference on Information & Communication Technologies, April 2013.
- [6] D. Deora and N. Bajaj, "Indian sign language recognition," 2012 1st International Conference on Emerging Technology Trends in Electronics, Communication & Networking, Dec 2012.
- [7] NR. Albelwi and Y. M. Alginahi, "Real-time arabic sign language (arsl) recognition," in International Conference on Communications and Information Technology, Jan 2012