Two-way Sign Language Interpreter

Harsh Shah, Maqsood Khan, Mukund Shivpuje, Santosh Dodamani

Department of Computer Engineering, University of Mumbai, Atharva College of Engineering, Malad, Mumbai, India

Publication Info

Article history:
Received : 12 February 2020
Accepted : 29 May 2020

Keywords:
ANN, CNN, Interpreter.

Corresponding author:
Harsh Shah
e-mail: *harsh143786.hs@gmail.com*

Abstract

A sign language translator is an important milestones in facilitating communication between the deaf community and the public. Sign language has a remarkable advancement that has grown over the past years. Unfortunately, there are some drawbacks that have come along with this language. Not everyone knows how to interpret a sign language when having a conversation with a deaf and dumb person. There is always a need to communicate using sign language. One finds it hard to communicate without an interpreter. To solve this, we need a product that is versatile and robust. We need to convert the sign language so that it is understood by common people and will help them to communicate without any barriers. .We hereby present the development and implementation of an Indian Sign Language (ISL) fingerspelling translator based on a convolutional neural network. We will divide our implementation in Two modules. The first module will handle the user input which will be taken by device Camera. Now the second module contains the pre-processing of the input images, the different sign of images will be identified with the help of CNN and then internally ANN will be applied on the processed image. The identified signed of image will be then compared with the Thousands of gesture data set stored and their associated output, which are already stored. Corresponding word will be shown to User. In the same way if User will give voice input then corresponding gesture will be shown to User. The main purpose of this project is to eliminate the barrier between the deaf and dumb and the rest.

1. INTRODUCTION

Machine learning provides a versatile and robust environment to work on. The machine learning subject also eliminates the need for the coder to write updates whenever a new sign is read, this will be done by the machine itself. Our system aims to get the deaf and dumb people more involved to communicate and the idea of a camera-based sign language recognition system that would be in use for converting sign language gestures to text (English) and then to regional languages. Our objective is to design a solution that is intuitive and simple which simplifies the communication for the majority of people with deaf and dumb' There are many methods to convert the sign language which often use Kinect as the basic system to get the inputs and work on them for conversion. Kinect methods are complicated in so many aspects. Our approach will be simple. We use simpler ways to capture the inputs and process them. We have used common and easily available libraries in our system.[1] The role of the interpreter appears to be very straightforward—to effectively facilitate communication between deaf individuals and those who are hearing. However, the complexities of the task, the varieties or types of visual interpreting, and the enormous range of qualifications brought by the interpreter make it anything but simple. Interpreting requires a high level of fluency in two or more languages, keen ability to focus on what is being said, broad-based world knowledge, and professional, ethical conduct.

Interpreters cannot interpret what they do not understand. Interpreters serve all parties in the communication exchange. Although we often think of the deaf person as the requestor of interpreter services, the reality is that all parties have an equal and mutual need for the interpreter. [2]

1.1. Need and Motivation

Sign languages aren't easy to learn and are even harder to teach. They use not just hand gestures but also mouthing, facial expressions and body posture to communicate meaning. This complexity means professional teaching programmes are still rare and often expensive. But this could all change soon, with a little help from artificial intelligence (AI). Practising and assessing sign language is hard because you can't read or write it. [3]

A mundane life of speech-impaired person is so hard that he/she can't do anything they would like as they could not communicate with common people. They become more dependent on people who translate for them. It would be easy if there is a Digital Assistant which fills the gap between the two by translating the gestures into sentences, which can be later transformed into voice to make the communication possible. Sign Language Project addresses this gap and acts as a bridge between speech-impaired and common people. The system can also be integrated with chat-bots and other voice assistants, so that they could use them without any hurdles. The study focuses on developing a digital assistant that recognizes the gestures (Sign Language) and translate them to proper sentences, which can be then used for enabling proper communication between speech-impaired and common people. Speech-impaired people feel difficult to communicate with common people through Sign Language in their regular routine. They understand only those gestures, which in turn makes it difficult for them to communicate effectively. The gestures vary between different countries and regions. A system that facilitates people to interpret the gestures irrespective of the region is to be built. [4]

1.2. Basic concept

Capturing signs from real world and translating them is the core objective of this work. The real-world signs are read using a webcam which captures both static and moving images of the objects in front of it. The deaf and dumb person who is signing is made to stand in front of the webcam and the image captured from this is processed with the tf-pose-estimation library to map out the skeleton of the person signing. Fig 2 is an example of how the skeleton is mapped on the system.

2. AIM AND OBJECTIVES

2.1. Aim

Our project's aim is to develop a two-way interpreter system which will use the concept of image processing to classify the sign from the given dataset. We intend to combine feature extraction technique and classification for the final output. To do so we are going to convert the variety of images into binary image during the pre-processing phase and then geometric features such as area will be calculated and condition will be applied. Though a lot of research has been done on quality and accuracy of the sign, we want to classify the interpretation and display the result which would help in reducing the human time and effort.

2.2. Objectives

1. The objective of our project is to develop two-way interpreter system which can accept a sign(movement) and classify the sign after evaluation of some intermediate phases.
2. Mathematical computations and storage of quality data would help in reducing the time
3. The system will classify the sign and the result will be made available to the user through the graphical user interface. The dataset of signs will be maintained in the folder with each image having a unique number.

3. PROPOSED SYSTEM

The idea is to facilitate the communication between hearing impaired and normal hearing people by building a translation system. The system will play the role of human sign language interpreter that understands the sign language and translates it into speech for normal hearing people. Three parties will be involved: the hearing-impaired person, system that consist a computer and normal hearing person. The hearing-impaired person performs signs in front of the System. System tracks the hearing-impaired person and transmits sign information to the computer. Computer analyses the sign and provides its meaning as speech to normal hearing person.

4. REQUIREMENT ANALYSIS (SRS)

4.1. Functional requirements

The functions that systems should provide to its user are known as functional requirements.

Although the fact that two people are involved, only one of them interacts directly with the system. This is the hearing-impaired person that performs sings in front of system and the system responds by translating them to speech. The other person can be considered as passive user since it does interact directly with the system, but only listens/reads to recognize signs.

Requirements can be represented by describing the interaction that different actors (users, external systems) have with the system:

- System is presented as black box.
- Hearing impaired is the person that performs the signs.
- Normal hearing is the passive user of the system.
- Therefore, the system requirements can be specified as follows:
- Hearing impaired person should be able to perform sign that represent digit number.
- Hearing impaired person should be able to perform sign that represent alphabet letter.

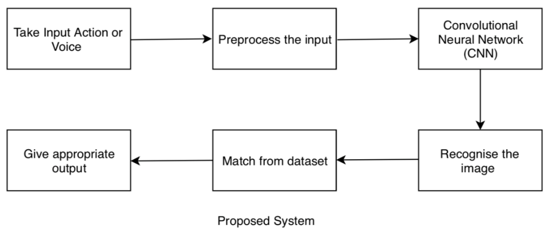

Figure 3.1: Proposed system

- Hearing impaired person should be able to perform sign that represent word.
- Hearing impaired person should be able to perform sign that represent sentence.
- Hearing impaired person should be able to see the translation of sign to text.
- Hearing impaired person should be able to change the component (number/alphabet or word/sentence) for which translation to speech is provided.

4.2. Non-functional requirements

The conditions on which system should operate are specified as non-functional requirements and they are:

1. Real time: the system should provide the recognition of signs and their translation to speech in an unnoticeable time for its users.
2. Accuracy: signs should not be confused and the system should recognize appropriate sign.
3. Environment: the system should provide real time recognition with high accuracy in low light conditions as well.
4. Usability: the system should provide natural interaction to its users. The hearing-impaired person needs to worry nothing else, just for performing signs.

5. SCOPE AND FEASIBILITY OF PROJECT

5.1. Scope

The current scenario in this field is that, lots of research have done and lots of white papers are there but there is no full-fledged Sign Language Interpreter which can translate signs to voice. This project is more inclined towards this particular side.

This project will take the input from camera and will convert it to voice and voice again back to sign, as required. Depending upon the dataset and the architecture is used we can expect the accuracy close to 90%. Impaired person when surrounded by the relatives or close people he will not face much problem to communicate for both parties. The problem arises when he wants to communicate with the person who is stranger. Solution provided by us will be helpful in every scenario.

5.2. Feasibility

5.2.1. Operational Feasibility

Operational Feasibility is the ability to utilize, support and perform the necessary tasks of a system such as storing of images and classifying it based upon the features. It includes everyone, from who creates, operates or uses the system. To be operationally feasible, the system must fulfil a need required by the business.

5.2.2. Technical Feasibility

Technical Feasibility, involves development of a working model of the product or a service. With developments in Neural networks, Artificial Intelligence, Deep leaning, the system can rely on such effective algorithms. This system depends upon the pool of Deep learning algorithms, which makes the system feasible.

6. DESIGN DETAILS

6.1. Context Level diagram

The Context Diagram shows the system under consideration as a single high-level process and then shows the relationship that the system has with other external entities (systems, organizational groups, external data stores, etc.).

Another name for a Context Diagram is a Context-Level Data-Flow Diagram or a Level-0 Data Flow Diagram.

6.2. Data Flow Diagram

A data flow diagram (DFD) maps out the flow of information for any process or system. It uses defined symbols like rectangles, circles and arrows, plus short text labels, to show data inputs, outputs, storage points and the routes between each destination. Data flowcharts can range from simple, even hand-drawn process overviews, to in-depth, multi-level DFDs that dig progressively deeper into how the data is handled. They can be used to analyse an existing system or model a new one.

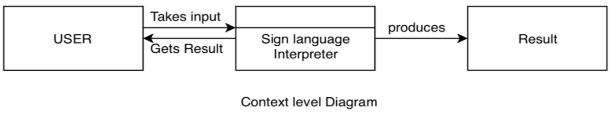

Figure 6.1: Context Level diagram

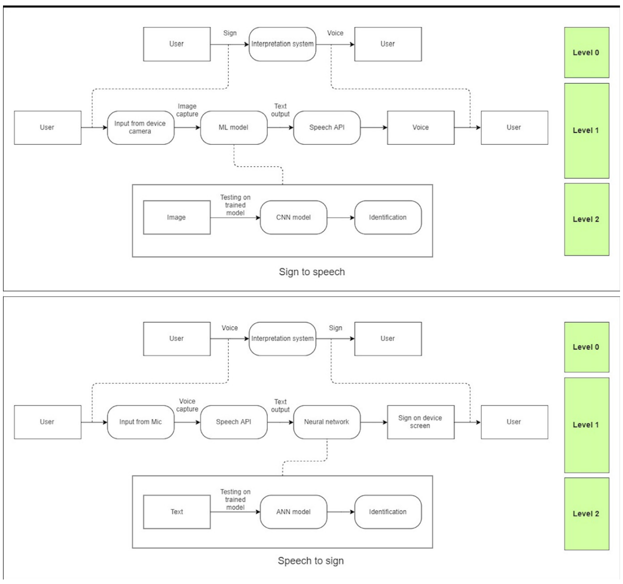

Figure 6.2: Data flow diagram

6.3. Sequence diagram

In the sequence diagram, the objects according to their time of execution are listed below. The sequence diagram deals with how the user will flow through the system sequentially and what are the processes that will take place. The objects interact with each other by sending messages to each other. As shown in the figure, the user first performs the sign gesture and then CNN internally perform many operations on the sign gesture then result will be made available to the user.

6.4. E-R Diagram

The main entities and how they are related with the other is shown in the diagram below. The entities and their key attributes are defined and what entities are interacting with each other for what purposes. There will be a dataset in which the collection of Sign images will be present and each Sign image will be given a unique number and this unique number will act as a primary key. Therefore, the Sign becomes a strong entity. The user selects a Sign and gives it to the Interpreter, the system will inspect the Sign for mapping it to corresponding text and so to voice.

6.5. Control Flow Diagram

The control flow diagram shows how the user will flow through the system, and how the user's data will flow. The diagram below shows how the user input will be converted to the output, and based on what the user wants to do. The diagram shows the decisions that the system will perform to get the desired output. The intermediate phases are shown in the diagram.

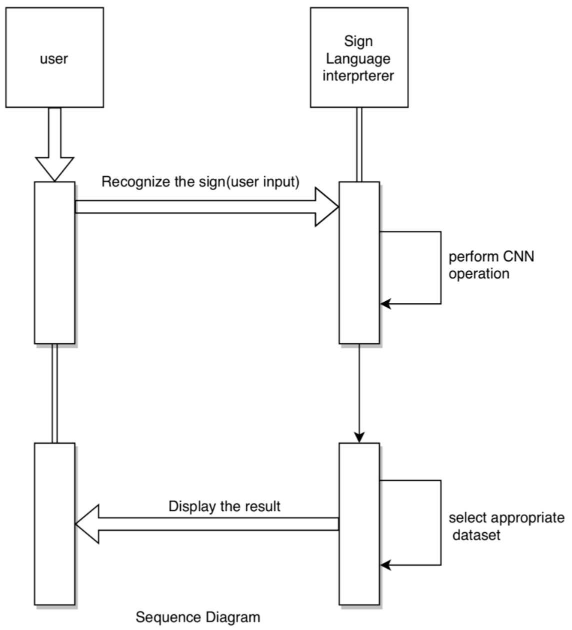

Figure 6.3: Sequence diagram

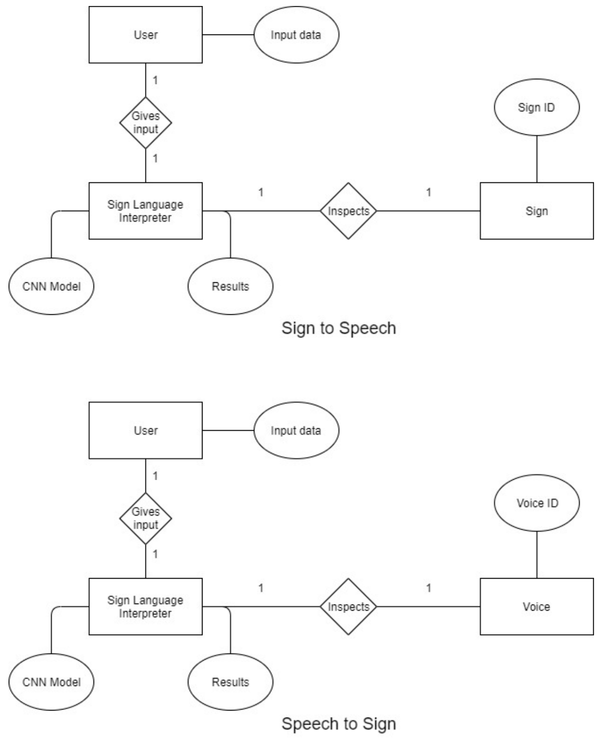

Figure 6.4: ER diagram

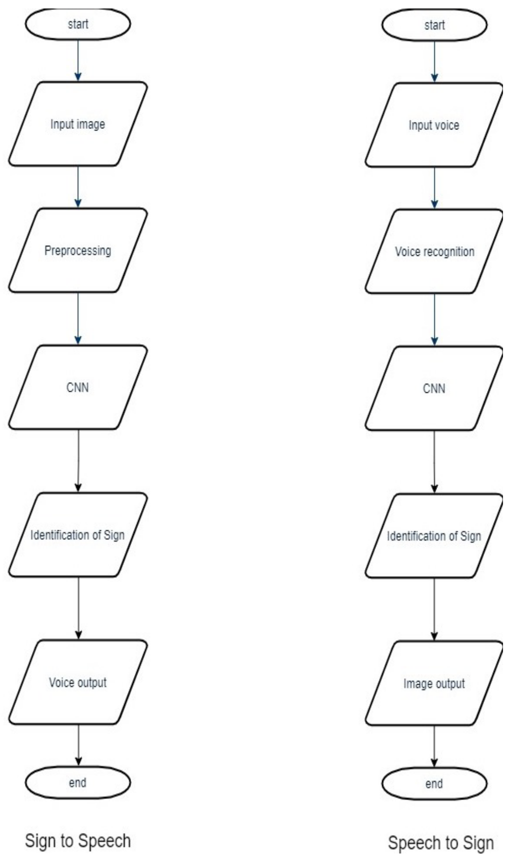

Figure 6.5: Control flow diagram

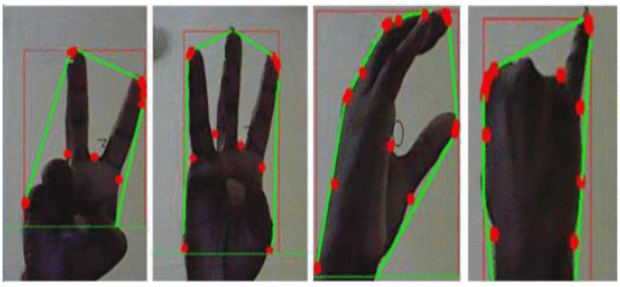

Figure 8.1: Input hand signs for number 2,3 and alphabet C, D

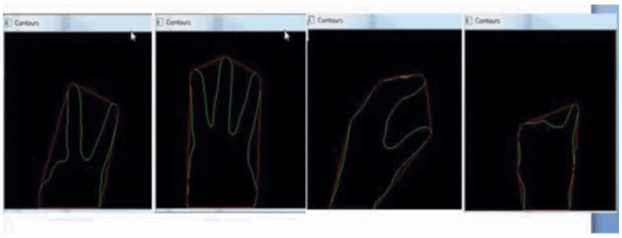

Figure 8.2: Contours of signs for numbers 2,3 and alphabets C, D

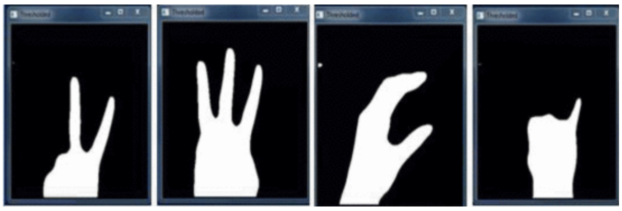

Figure 8.3: Threshold images of numbers 2,3 and alphabets C, D

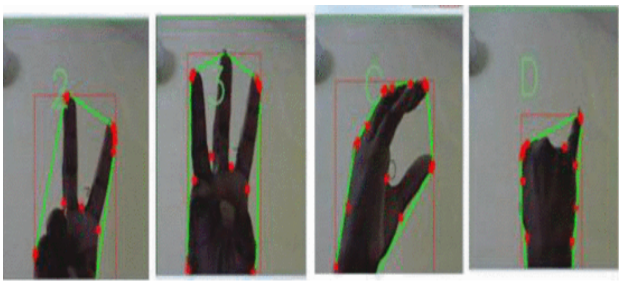

Figure 8.4: Output images with text Display T of numbers 2,3 and alphabets C, D

7. METHODOLOGY

7.1. Building Architecture

Our work over the past months focused on the design and training of neural networks that effectively use our growing dataset. We studied several architectures to come up with a solution that both meets our high-performance requirements and creates minimal runtime overhead. In the end we converged on an architecture that contains a three-dimensional convolutional network (3D-CNN) to extract spatiotemporal features, a recurrent layer (LSTM) to model longer temporal relations, and a SoftMax layer that outputs class probabilities[9].

7.2. Preparing Dataset

Preparing dataset was rather a lengthy process.

The videos clips are challenging, because they capture the complex dynamics of the real world[7].

While the gesture is easy to recognize for humans, it is difficult to understand for a computer because the video footage contains sub-optimal lighting conditions and background noise.

7.3. Training

Training such type of complex model is a difficult task altogether

After researching and trying all of the possible information we have come to the conclusion that training in more passes, that is, epochs overfits the model[10].

Also, there should be distinct non overlapped data elements in the train and test subsets, so as to make the model more robust. In the training data there should be different classes to make model more diverse[8].

8. IMPLEMENTATION AND RESULT

The python programming is used to implement the system. The input to the system is shown in Fig 11.1. The symbols for numbers 2 and 3 and alphabets c and d are shown in Fig 11.1. The contour of the images is then obtained which is shown in Fig 11.2. The binary images are shown in Fig 11.3 this binary or threshold images are compared with the template images stored in database. After the match is performed corresponding text is obtained for each hand sign. This is shown in the output figure Fig 8.4.

9. CONCLUSION

Manual signs are composition of hand shape, location, orientation, and movement. The developed system incorporates these components and uses them to recognize and translate to speech: numbers, alphabet letters, words and sentences from Sign Language. In general, the system can achieve higher recognition accuracy for one hand signs and lower recognition accuracy for signs where both hands were involved. While the recognition accuracy for static signs was high (around 90%), further work is needed in order to improve recognition accuracy of hand shape component when it is combined with hand movement and hand location features[7]. A system that provides complete Sign Language Recognition is not done yet to the best of author knowledge and as presented in related work section, research is being done continuously toward building these types of systems.

10. ACKNOWLEDGEMENTS

We would like to thank our project guide Prof. Santosh Dodamani for his enormous cooperation and guidance. We have no words to express our gratitude for a person who wholeheartedly supported the project and gave freely of his valuable time while making this project. All the inputs given by him found a place in the project. The technical guidance provided by him was more than useful and made this a success. He has always been a source of inspiration for us. It was a great experience learning under such a highly innovative, enthusiastic and hardworking professor. We are also thankful to our principal Dr.S.P.Kallurkar, and HOD of Computer Department, Prof. Suvarna Pansambal, project coordinators, and all the staff members of the Computer Department who have provided us with various facilities and guided us throughout to develop this project idea. Finally, we would like to thank teachers of the college and friends who guided and helped us while we worked on this project idea.

11. REFERENCE

[1] H.G.M.V.C.H.N.T. Vishwas S, "SIGN LANGUAGE TRANSLATOR USING MACHINE LEARNING," Research India Publications, 2018.

[2] NDC, "Sign Language Interpreters: An Introduction," NDC, Austin, 2019.

[3] M. Ketchell, "How AI could help you learn sign language," The Conversation, p. 1, 17 January 2019.

[4] "Sign Language Recognition using Depth Data and CNN," SSRG International Journal of Computer Science and Engineering, Trichy, 2019.

[5] E Hunter, J Schlenzig, R Jain, "(1995) Posture Estimation in Reduced-Model Gesture Input System", *Int. Workshop on Automatic Face and Gesture Recognition*, pp. 290-295, 26-28 June 1995.

[6] D Shridhar, A Badreldin, "High-Accuracy Character Recognition Algorithm using Fourier and Topology Descriptors", *Pattern Recogn*, vol. 17, pp. 515-524, 1984.

[7] Lopez, S. Rio, J.M. Benitez and F. Herrera, "Real-time sign language recognition using a consumer depth camera", IEEE International Conference on Computer Vision Workshops,2013.

[8] T. Watanabe and M. Yachida " Real Time Gesture Recognition Using Eigenspace from Multiple Input Image Sequences" Proc. of FG'98 pp. 428-433 1998.

[9] W. Stokoe, D. Casterline, and C. Croneberg, A Dictionary of American Sign Language on Linguistic Principles, Gallaudet College Press, Washington D.C., USA, 1965.

[10] D. Keysers, T. Deselaers, C. Gollan, and H. Ney, "Deformation models for image recognition," IEEE Trans. PAMI, p. to appear, 2007.