**“Gesture Based Real-time Indian Sign Language Interpreter”**

Submitted in partial fulfillment of the requirements

of the degree of

Bachelor of Computer Engineering

by

**Akshay Divkar (BE A 75)**

**Rushikesh Bailkar (BE A 04)**

Under Guidance of

**Dr. Chhaya S. Pawar**



**University of Mumbai**

**Department of Computer Engineering**

**Datta Meghe College of Engineering**

Plot no.98 Sector-3, Airoli, Navi Mumbai

**Academic Year 2020-21**



**Datta Meghe College of Engineering**

**Airoli, Navi Mumbai**

**CERTIFICATE**

This is to certify that the project entitled “**Gesture Based Real-time Indian Sign Language Interpreter**” is bona fide work of “**Akshay Divkar, Rushikesh Bailkar**” submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of “Undergraduate” in “Computer Engineering”.

Dr. C. S. Pawar Dr. A. P. Pande Dr. S. D. Sawarkar Project Guide Head of the Department principal



**Datta Meghe College of Engineering**

**Airoli, Navi Mumbai**

**Project Approval**

This project report entitled “Gesture Based Real-time Indian Sign Language Interpreter” of the students “Akshay Divkar, Rushikesh Bailkar” approved for the degree of Computer Engineering.

Internal Examiner External Examiner

Date: Date:

Place: Place:

**Declaration**

We declare that, this written submission represents our ideas in our own words and where others' ideas or words have been included; we have adequately cited and referenced the original sources. We also declare that, we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Signature

Akshay Divkar

Rushikesh Bailkar

**Acknowledgement**

Motivation and guidance are the keys towards success. I would like to extend my thanks to all the sources of motivation.

We would like to grab this opportunity to thank **Dr. S. D. Sawarkar,** principal for encouragement and support he has given for our project.

We express our deep gratitude to **Dr. A. P. Pande**, head of the department who has been the constant driving force behind the completion of this project.

We wish to express our heartfelt appreciation and deep sense of gratitude to my project guide **Dr. C. S. Pawar** for her encouragement, invaluable support, timely help, lucid suggestions and excellent guidance which helped us to understand and achieve the project goal. Her concrete directions and critical views have greatly helped us in successful completion of this work.

We extend our sincere appreciation to all professors for their valuable inside and tip during the designing of the project. Their contributions have been valuable in so many ways that we find it difficult to acknowledge of them individually.

We are also thankful to all those who helped us directly or indirectly in completion of this work.

Place:

Date:

**Abstract**

Hand gesture is one of the methods used in sign language for non-verbal communication. It is most commonly used by hearing & speech impaired people who have hearing or speech problems to communicate among themselves or with normal people. Developing sign language application for hearing impaired people can be very important, as hearing & speech impaired people will be able to communicate easily with even those who don’t understand sign language. This project aims at taking the basic step in bridging the communication gap between normal people, deaf and dumb people using sign language. The main focus of this work is to create a vision based system to identify sign language gestures from the video sequences. The reason for choosing a system based on vision relates to the fact that it provides a simpler and more intuitive way of communication between a human and a computer. Video sequences contain both the temporal as well as the spatial features. In this project two different models are used to train both the temporal as well as the spatial features. To train the model on the spatial features of the video sequences a deep Convolutional Neural Network. Convolutional Neural Network was trained on the frames obtained from the video sequences of train data. To train the model on the temporal features Recurrent Neural Network is used. Trained Convolutional Neural Network model was used to make predictions for individual frames to obtain a sequence of predictions. Now this sequence of prediction outputs was given to Recurrent Neural Network to train on the temporal features. Collectively both the trained models i.e. Convolutional Neural Network and Recurrent Neural Network will produce the text output of the respective gesture.

**INDEX**

Acknowledgement i

Abstract ii

List of Figures iv

List of Nomenclature v

|  |  |  |  |
| --- | --- | --- | --- |
| 1. | Introduction | | 1 |
|  | 1.1 Background  1.2 Sign Language  1.3 Motivation |  | 1  2  2 |
| 2. | Literature Survey | | 3 |
|  | 2.1 Introduction  2.1.1 Different Implementations of Sign Language Recognition  2.1.1.1 Hardware Based  2.1.1.2 Gesture Type Based  2.2 Previous Work on SLR  2.3 Findings of literature Review |  | 3  4  4  5  7  9 |
| 3. | Requirement Analysis | | 11 |
|  | 3.1 Functional Requirement  3.2 Non-Functional Requirement  3.3 Software & Hardware Requirement |  | 11  12  13 |
| 4. | Proposed System | | 14 |
|  | 4.1 Problem Definition  4.2 System architecture  4.2.1 Video acquisition & Pre-processing  4.2.2 Object localization  4.2.3 Feature extraction  4.3 Objectives |  | 14  14  15  16  16  18 |
| 5. | Project Design | | 19 |
|  | 5.1 Data Flow Diagram  5.1.1 Level 0 Data Flow Diagram  5.1.2 Level 1 Data Flow Diagram |  | 19  19  20 |
| 6. | Technology Used | | 21 |
|  | 6.1 CNN Model  6.2 RNN Model |  | 21  21 |
| 7. | Conclusion and Future Scope | | 23 |
|  | 7.1 Conclusion  7.2 Future Scope |  | 23  23 |
|  | References | | 24 |

**List of Figures**

|  |  |  |
| --- | --- | --- |
| Sr. No. | Name of the Figure | Page No. |
| 2.1 | Sign Language Recognition Approaches | 4 |
| 2.2 | Gesture Types | 5 |
| 2.3 | Static Gesture example | 6 |
| 2.4 | Dynamic Gesture Example | 6 |
| 4.1 | System Architecture SLR | 15 |
| 4.2 | Convolution Neural Network | 17 |
| 5.1 | DFD Level 0 Sign Language Recognition | 19 |
| 5.2 | DFD Level 1 Sign Language Recognition | 20 |
| 6.1 | RNN repetitive block | 22 |

**List of Nomenclature**

|  |
| --- |
| SLR – Sign Language Recognition |
| CNN - Convolutional Neural Network |
| RNN – Recurrent Neural Network |
| ISL – Indian Sign Language |
| ASL – American Sign Language |

**CHAPTER 1**

**INTRODUCTION**

Gestures are naturally performed by humans. Gestures are produced as part of deliberate actions, signs or signals, or subconsciously revealing intentions or attitude. They may involve the motion of all parts of the body, but the arms and hands, which are essential for action and communication, are often the focus of studies. Facial expressions are also considered gestures and provide important role in communication. Gestures are present in most daily human actions or activities, and participate to human communication by either complementing speech or substituting themselves to spoken language in environments requiring silent communication (underwater, noisy environments, secret communication, etc.) or for people with hearing disabilities.

**1.1 Background**

The goal of this project is to build a neural network able to classify which word/sentence of the Indian Sign Language (ISL) gesture is being signed. This project is a first step towards building a possible sign language interpreter, which can take interpreter, would greatly lower the barrier for many hearing impaired and speech impaired individuals to be able to better communicate with others in day to day interactions. In India, there is no universal sign language. Even though there exist many sign languages, the normal people do not know about sign languages. Gesture recognition and sign language recognition has been a well-researched topic for American Sign Language but has been rarely touched for its Indian communications in sign language and translate them into written language. Such an counterpart. Hence communicating with hearing impaired and speech impaired people become more complex. Most research implementations for this task have used depth maps generated by depth cameras and high resolution images. The objective of this project is to see if neural networks are able to classify signed ISL words using sequences of images of hands taken with a personal device such as a laptop webcam. This is in alignment with the motivation as this would make a future implementation of a real time ISL-to-oral/written language translator practical in an everyday situation

**1.2 Sign Language**

Sign language is a communication medium used by hearing impaired and speech impaired people to exchange information between their own community and with other people. Computer recognition of sign language deals from sign gesture acquisition and continues till text/speech generation. Sign gestures can be classified as static and dynamic. However static gesture recognition is simpler than dynamic gesture recognition but both recognition systems are important to the human community.

**1.3 Motivation**

While automatic speech recognition has now advanced to the point of being commercially available, automatic SLR is still in its infancy. Currently all commercial translation services are human based, and therefore expensive, due to the experienced personnel required. SLR aims to develop algorithms and methods to correctly identify a sequence of produced signs and to understand their meaning. There are mainly two different motivations for developing sign language recognition model. The first aspect is the development of an assistive system for the hearing impaired people. Moreover hearing people have difficulties in learning sign language and likewise the majority of those people who were born deaf or who became deaf early in life, have only a limited vocabulary of accordant spoken language of the community in which they live. Hence a system of translating sign language to spoken language would be of great help for deaf as well as for hearing people. A second aspect is that sign language recognition serves as a good basis for the development of gestural human-machine interfaces.

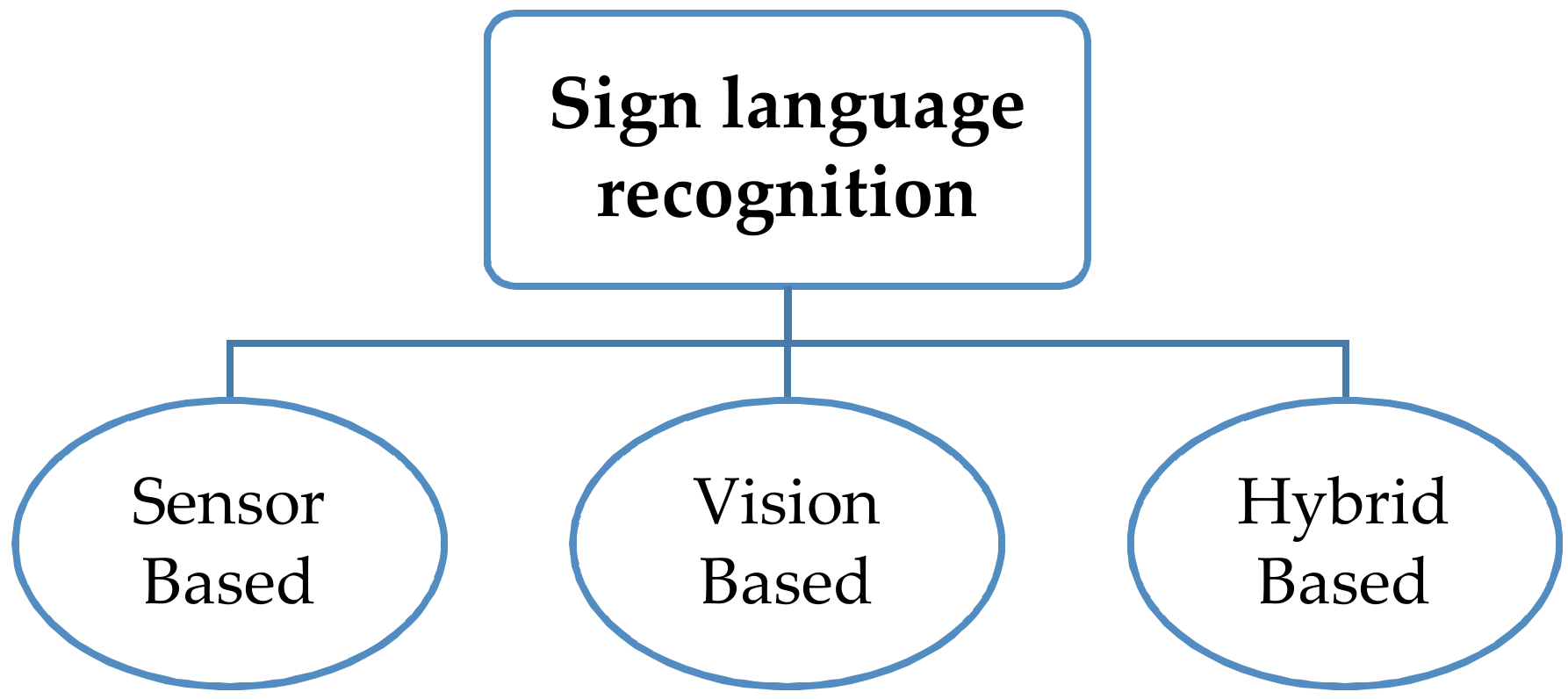
**CHAPTER 2**

**LITERATURE SURVEY**

A Literature Survey is a search and evaluation of the available literature in the topic area. It documents the state of the art with respect to the topic. Literature Surveys are secondary sources and do not report new or original experimental work. Literature Survey is conducted to identify where and how new research fits into the existing body of literature in a particular field of study. It highlights the strengths and weaknesses of previous research on a topic and makes recommendations for further research.

**2.1 Introduction**

Sign Language Recognition Application Systems is developed in two steps, data acquisition and classification. There are two data acquisition methods that are often used by researchers, camera and Microsoft Kinect. Some uses camera for their Sign Language Recognition Systems. The main advantage from using camera is that it removes the needs of sensors in sensory gloves and reduces cost from building the system. The camera is quite cheap and is available in almost all laptops. Some system uses high specification camera because of the blur caused by web camera. But even though it is high specification camera, it is still available in most of smartphones. High specification cameras are used to acquire the detailed data they need. The disadvantage of using web camera, or simply camera, is that good image pre-processing of obtaining the feature is needed. The Microsoft Kinect is the other popular method used by researchers to acquire their data. Microsoft Kinect is getting more popular among researchers as it provides more data and it is needed by researchers. Kinect sensor gives image with depth to acquire their data. The advantage of using Kinect is that it provides the depth data of the video stream. The depth data is very useful as it can easily distinguish the background and the signer. Furthermore, it can be used to distinguish hands and body as the signer usually performs sign language by hands in front of their body. The disadvantage is that the Microsoft Kinect device is costly and it should be connected to computer. Another technique of simple camera and color gloves to differentiate both hands and ease the feature extraction process. Glows are using 3-axis accelerometer and flex sensors. All of these gloves are equipped with sensors attached to the gloves. The advantage is that it provides all the data needed more accurately as it also provides fingers movement data. The disadvantages are that they are costly and are difficult to be used commercially. There are many existing systems of SLR most of which are based on static gesture recognition for various spoken languages and there are very few which are based on dynamic gestures but only for American Sign Language.



**Fig 2.1:** Sign Language Recognition Approaches

**2.1.1 Different Implementations of Sign Language Recognition**

There are different implementations of Sign Language recognition. Which are based on hardware used and the type of image sensor are used. The use of different hardware implementation results in different system architecture. So the methodologies are also changed.

**2.1.1.1 Hardware Based**

1. Glows based (Sensor Based)

In the glove based system, sensors such as potentiometer, accelerometers etc. are attached to each of the fingers. Based on their readings the corresponding alphabet is displayed. It is expensive to implement, not handy to use and requires mapping every time new user wear it. A major advantage of glove-based systems over vision-based systems is that gloves can directly report relevant and required data (degree of bend, pitch, etc.) in terms of voltage values to the computing device, thus eliminating the need to process raw data into meaningful values.

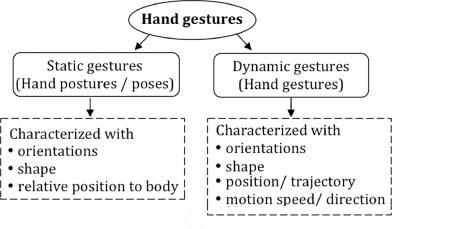
1. Image sensor based (Vision based)

Vision-based systems use cameras as primary tools to obtain the necessary input data. The main advantage of using a camera is that it removes the need for sensors in sensory gloves and reduces the building costs of the system. Cameras are quite cheap, and most laptops use a high specification camera because of the blur caused by a web camera. A well implemented system can give good result.

1. Gloves and Image sensor based (Hybrid based)

The third method of collecting raw gesture data employs a hybrid approach that combines gloves and camera-based systems. This approach uses mutual error elimination to enhance the overall accuracy and precision. However, not much work has been carried out in this direction due to the cost and computational overheads of the entire setup. Nevertheless, augmented reality systems produce promising results when used with hybrid tracking methodology

**2.1.1.2 Gesture Type Based**



**Fig 2.2:** Gesture Types

1. Single Point Gesture Recognition (static)

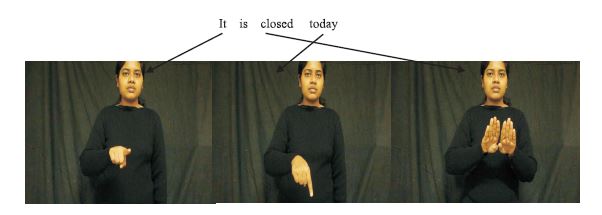
Static gestures are those that only require the processing of a single image at the input of the classifier, the advantage of this approach is the lower computational cost. It is referred as 2D gesture recognition. Working of static gesture based system is shown in **Fig 2.3**

****

**Fig 2.3:** Static Gesture example

1. Dynamic Gesture Detection

Dynamic gestures require the processing of image sequences and more complex gesture recognition approaches. It is referred to as 3D gesture recognition. Working of dynamic gesture based system is shown in Fig 2.4



**Fig 2.4:** Dynamic Gesture Example

**2.2 Previous Work on SLR**

This section contains the information about current available approaches of Sign Language Recognition methods based on the evaluating parameters (Method used, Functioning, Obtained results). In addition, this also evaluates the performance and limitations of available methods.

**2.2.1 P. V. V. Kishore, D. A. Kumar, A. S. C. S. Sastry and E. K. Kumar, "Motionlets Matching With Adaptive Kernels for 3-D Indian Sign Language Recognition," in IEEE Sensors Journal, vol. 18, no. 8, pp. 3327-3337, 15 April15, 2018, doi: 10.1109/JSEN.2018.2810449.**

A model for recognizing gestures of Indian sign language 3D motion captured data is presented. The model builds a two phase algorithm which handles multiple attributes of 3D sign language motion data for machine translation. In phase–I, the unordered 3D sign database is restructured into a 4–class structured motionlet database from the measured trajectories of motion segmented 3D joints. Each action in a signed frame is motion segmented into motion joints and non-motion joints. Phase–II extracts shape and orientation of 3D motionlets by applying joint relative distance and joint angle measurements respectively. Three feature kernels based on trajectories, finger shape and their orientations are constructed, which measure the similarity between the query signs and the database signs. It is observed that the motionlet based adaptive kernel matching algorithm on 500 class 3D sign language data gives better classification accuracies compared to state–of–the–art action recognition models. More importantly, it significantly optimizes the database search space by 75% over the existing kernel matching methods. The system has gained accuracy of 98.9 %.

**2.2.2 H. Muthu Mariappan and V. Gomathi, "Real-Time Recognition of Indian Sign Language," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1-6, doi: 10.1109/ICCIDS.2019.8862125.**

The system for recognizing real-time Indian Sign Language (ISL) portrays an impressive role in enhancing casual communication among people with hearing disabilities and normal persons. Though FCM is efficient, it requires more computation time than the others. Also, for high dimensionality datasets, most of the traditional algorithms suffer. Hence it is planned to extend the system by combining Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) to capture the spatial and temporal features. In future work, more words will be added to the system. This FCM based real-time sign language recognition system, for recognizing the words of Indian Sign Language has produced 75 % accuracy in gesture labeling and this is somewhat higher than the similar systems. Also, the developed system is much better than other systems, since it is capable of recognizing 40 words of ISL in real-time while the similar systems have the capability to recognize static gestures only. The FCM is more efficient and reliable.

**2.2.3 T. Oliveira, N. Escudeiro, P. Escudeiro, E. Rocha and F. M. Barbosa, "The VirtualSign Channel for the Communication Between Deaf and Hearing Users," in IEEE Revista Iberoamericana de Tecnologias del Aprendizaje, vol. 14, no. 4, pp. 188-195, Nov. 2019, doi: 10.1109/RITA.2019.2952270.**

This article details the improvements and current structure of the VirtualSign platform, a bidirectional sign language to text translation tool in development. The platform has two main components, sign to text and text to sign, that are both described. Translation from text to sign relies on a 3D avatar. Translation from sign to text relies on a set of data gloves. Two different data glows are used for testing, which are 5TD Data Glows and IFG Data Glows. The translator consist two modes of workings, the “word mode”, for the regular translation defined on previous paragraphs, and the “spelling mode”. This mode is activated automatically when the user performs three distinct alphabet hand gesture. Results of the system show that the 5DT data gloves perform significantly better than the IFG data gloves. On the 20, 50 and 100 words dataset, the 5DT system achieved 86%, 78% and 70% accuracy respectively when considering the two most likely signs. The IFG system performance is significantly worth, achieving an accuracy of 40%, 26% and 18% respectively.

**2.3 Findings of literature Review**

There is No dataset available for dynamic recognition system for Indian Sign Language. In India, there is no universal sign language. Though there exists many Sign Languages, the normal people do not know about sign languages. Gesture recognition and sign language recognition has been a well-researched topic for American Sign Language but has been rarely touched for its Indian equivalent.

By the conducted survey it is found that there is no such a system for Indian Sign Language which can give output as complete word or phrase for a single motion gesture.

**Table 2.1:** Literature survey

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Year** | **Paper** | **Author** | **Advantages** | **Disadvantages** |
| 2018 | Motionlets Matching With Adaptive Kernels for 3-D Indian Sign Language Recognition | P. V. V. Kishore, D. A. Kumar, A. S. C. S. Sastry and E. K. Kumar | 1) Good accuracy.  2) makes use algorithm only | 1) Much expensive  2) Large in physical size, Not portable  3) Used 9 3d camera  4) complex implementation |
| 2019 | Real-Time Recognition of Indian Sign Language | Muthu Mariappan H,  Dr Gomathi V | 1) Less expensive  2) implemented using camera | 1) Used static gestures. |
| 2019 | The Virtual Sign Channel for the Communication  Between Deaf and Hearing Users | Tiago Oliveira , Nuno Escudeiro Paula Escudeiro Emanuel Rocha | 1) No need for training dataset | 1) based on Data glows  2)expensive because  of glows |

There are several approaches that have been used to design a hand gesture recognition system. The main three different approaches are given in the Table 2.1 with their advantages and disadvantages In all the approaches, the primary focus was given to feature extraction of the hand gesture and it was found that the better feature extraction step is performed, better will be the performance of classification. P.V.V. Kishore, proposed a method of recognizing hand gesture, by considering 3D skeleton model of the upper human body as a feature [1]. Which captures 3D motion of human joint positions and skeleton but a 3D skeleton model has no information about face and images. So this paper misses on the some key features to completely recognize sign language. Muthu Mariappan H and Dr Gomathi V, have recognized user hand gesture using FCM algorithm as their feature, in which, The features such as number of points in the convex hull, number of defect points and distance from the Centre to each finger are extracted from the Regions of Interest, through these three contours, and the orientation between the contours is also kept track [2]. The system is based on static gesture. Though FCM is efficient, it requires more computation time than the others. Tiago Oliveira, came up with the system based on Hand glows or data glows [3]. The system based on data glows only operate on the hand coordinates or the gesture coordinates. He observed that the finger-tip can be used as a feature for recognition and misses on the face data.

**CHAPTER 3**

**REQUIREMENT ANALYSIS**

Requirements Analysis is the process of defining the expectations of the users for an application that is to be built or modified. It involves all the tasks that are conducted to identify the needs of different stakeholders. Requirement analysis is an essential part of any project and project management. Without this, it is virtually impossible to successfully meet project goals. Therefore requirements analysis means to analyze, document, validate and manage software or system requirements. High-quality requirements are documented, actionable, measurable, testable, traceable, helps to identify business opportunities, and are defined to a facilitate system design.

**3.1 Functional requirement**

Functional requirements define the basic system behavior. Essentially, they are what the system does or must not do, and can be thought of in terms of how the system responds to inputs. Functional requirements usually define if/then behaviors and include calculations, data input, and business processes.

The functional requirements for Sign Language recognition system are as listed below.

i. The system should be able to receive a Video sequence through image sensor as input.

ii. It needs to determine whether the gesture is directed to the image sensor

iii. There should be ample of light in the performing environment.

iv. The performed gesture should have some meaning in the trained model.

v. It should give proper output respective to the input gesture.

vi. The working of system should be naturally aspired and user friendly.

**3.2 Non-Functional Requirement**

In systems engineering and requirements engineering, a non-functional requirement is a requirement that specifies criteria that can be used to judge the operation of a system, rather than specific behaviors. They are contrasted with functional requirements that define specific behavior or functions.

**3.2.1 Performance**

The respective output of the gesture should not take time more than 1-3 seconds.

**3.2.2 Reliability**

The output text of the system should be respective to the input gesture 98% of the time.

**3.2.3 Availability**

System should be available throughout the week at any time during the day. In the case of unplanned system downtime, all features will be available again after one working day.

**3.2.4 Maintainability**

The maintaining of the system should be possible.

**3.2.5 Recoverability**

If a major incident happens with the system must take measures to go back to being fully operational within three days.

**3.2.6 Capacity**

Till now system can handle communication between one signer and many listeners.

**3.2.7 Serviceability**

The addition of the new gestures can be added and replaced by uploading data and training model; there’s no need to recompile any code.

**3.2.8 Security**

Only the users with the role “developers” can make changes and addition to the system.

**3.2.9 Data integrity:**

The system shall maintain data integrity by keeping backups of all updates to the database for trained model.

**3.2.10 Usability**

The system’s interface has to be user-friendly and easy to use.

**3.3 Software & Hardware Requirement**

* Desktop / Laptop Processor : Intel i3 Core processor & Above
* Desktop / Laptop RAM : 4GB & Above
* GPU : 1GB & clocked at 900 MHz and above
* Hard Disk : 500GB & Above
* Monitor : Desktop / Laptop Monitor
* Mouse : Two or Three button mouse
* External Camera :USB attachable external camera
* Desktop / Laptop O. S : Windows 10 &Above, Ubuntu
* Software Environment : MS Visual Studio code

**CHAPTER 4**

**PROPOSED SYSTEM**

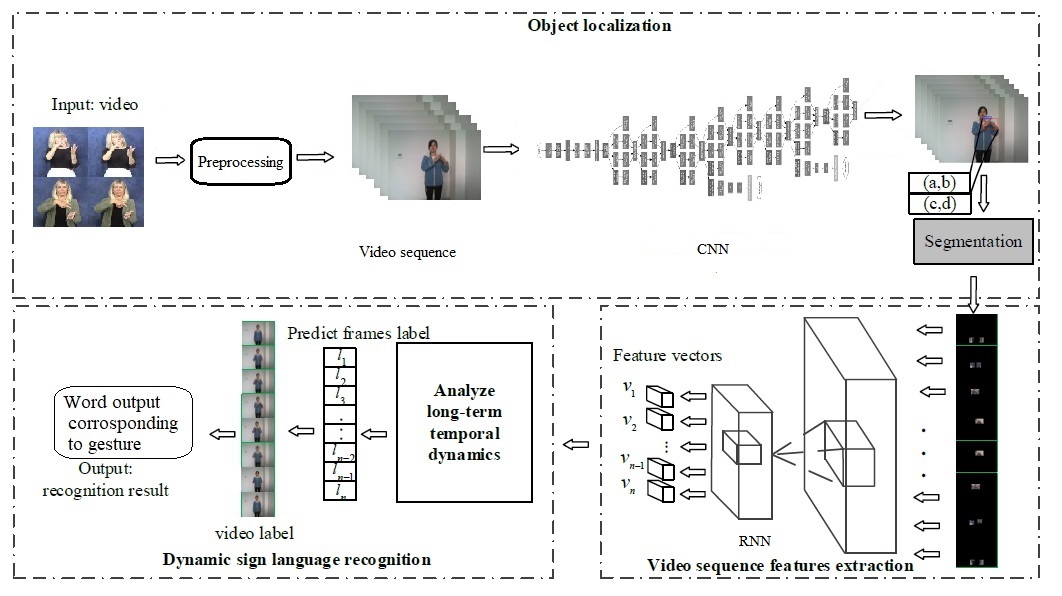
Technology is very fast growing and incredible, yet there are not much technology development and improvement for speech and hearing impaired peoples. Existing mobile applications and console gaming use sign language as the only option for communication and interaction with them. Since the sign language recognition is not that advanced speech and hearing impaired people are left behind from the society. Therefore Sign Language Recognition can help to bridge this gap.

**4.1 Problem Definition**

The persistent problem in the present Indian Sign Language Recognition is that all the implemented systems rely on static gesture recognition which is very slow or not handy to be used by speech impaired people. In all the existing systems they use single gesture for single character which takes much time and isn't efficient working of system.Speech impaired people use hand signs to communicate, hence normal people face problem in recognizing their language by signs made. Hence there is a need for systems which recognize the different signs and conveys the information to the normal people. Aim of this project is to develop a concept of virtual talking system without sensor for people who in need, this concept achieving a by using image processing and human hand gesture input. This mainly helps people who can’t talk with other people.

**4.2 System Architecture**

An architectural diagram is a diagram of a system that is used to abstract the overall outline of the software system and the relationships, constraints, and boundaries between components. It is an important tool as it provides an overall view of the physical deployment of the software system and its evolution roadmap. Fig 4.1 shows the architecture diagram of Sign language recognition.

****

**Fig 4.1:** System Architecture of SLR

**4.2.1 Video acquisition & Pre-processing**

Image hand gesture acquisition, as illustrated by Microsoft (2019) is to capture the human hand gesture image by the computer. This could be done using vision-based recognition where no special gadgets are required and a web camera or a depth camera is used, furthermore special tools can be utilized such as wired or wireless gloves that detect the movements of the user hand, and motion sensing input devices like (Kinect from Microsoft, PC Webcam, Leap Motion, etc.). The system will include low-resolution web cam for capturing the hand gestures and an algorithm that processes the acquired images and forwards the set of frames to the Pre-processing. Pre-processing only extracts the hand gesture from an image by selecting the Region of interest. Pre-processing also consist removal of the background noises (if present) and unwanted region. Pre-processing stage involves Skin Filtering which only selects the skin and drops rest of the image after which Feature vector based Feature Extraction takes place. As per the requirements for feature extraction the video sequence frames undergo resizing and normalization. The output from this is carry forward to feature extraction.

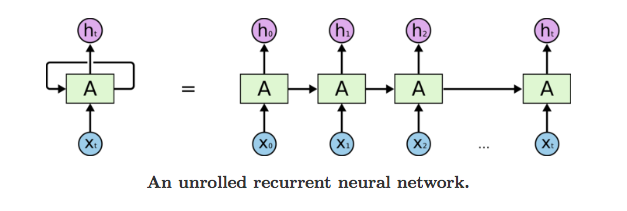
**4.2.2 Object localization**

Hand detection has great importance in temporal segmentation and the subsequent recognition module. In order to obtain the accurate information of hand location in the frame images, it’s essential to choose excellent object detection algorithm. Object localization module based on CNN, which is used to capture the information of hand position. The video frames are trained by Convolution layers for feature extraction. Convolutional Neural Networks allow for more efficient computation and deeper Networks through a dimensionality reduction. As a neural network deals with a vast array of images, with wide variation in the featured image content, they need to be designed appropriately. Each layer of CNN will work on individual Frames from the segmented video. In this every layer of CNN will perform Resize, Normalize, Image Segmentation, Hand Detection, and Feature Extraction. CNN will compute spatial data from the given input of video frames, this data is forwarded to RNN.

**4.2.3 Feature extraction**

Hand tracking process is the ability of the computer to trace the user hand and separate it from the background and from the surrounding objects. This can be done using multi-scale color feature hierarchies that gives the users hand and the background different shades of colors to be able to identify and remove the background, or by using clustering algorithms that are capable of treating each finger as a cluster and removing the empty spaces in-between them. Feature extraction methods are used in order to extract the useful information from the images that helps in the recognition process of gestures. These features can be extracted using several methods available, such as Fourier descriptor method which captures the palm, the fingers and the fingertips, or centroid method which captures the essential structure of the hand. The features extracted change from one application to another, some of the features that could be taken into consideration are: fingers status, thumb status, skin color, alignments of fingers, and the palm position

RNN is used for spatiotemporal feature extraction. Recurrent Neural Network is a generalization of feed forward neural network that has an internal memory. RNN is recurrent in nature as it performs the same function for every input of data while the output of the current input depends on the past one computation as shown in Fig 4.2. After producing the output, it is copied and sent back into the recurrent network. For making a decision, it considers the current input and the output that it has learned from the previous input. In other neural networks, all the inputs are independent of each other. But in RNN, all the inputs are related to each other. In this system, multiple layers of CNN are connected to RNN. RNN has the spatial output opted from the CNN layers. Collectively spatial output forms motion of the gesture and forms the spatiotemporal input for the RNN.

****

**Fig 4.2:** Convolution Neural Network

The classifier model provides facility to create a joint representation for these independent streams. It can analyze long-term temporal dynamics and predict the hand gesture label. Through analyzing each video feature vectors, the frames label could be predicted. Thus, the video sequence label could be predicted.

**4.3 Objectives**

* To develop a project that would Speech impaired people to get more involved in normal society.
* Implementing a single gesture for a word and single gesture for phrase so that it will be more natural.
* The problem statement revolves around the idea of a camera­ based sign language recognition system, so that the end product will be Cost efficient.
* Objective of this project is to design a solution that is intuitive, simple and user friendly.
* Communication for the normal society is not difficult. It should be the same way for the Speech impaired individuals.

**CHAPTER 5**

**PROJECT DESIGN**

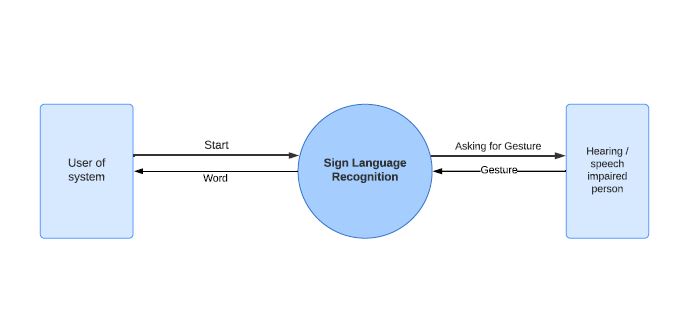
The lots of task are carried out to design a system that will correctly classify and give the output for its corresponding input. Design Systems are the single source of truth for the product experience. An ideal design system is built with reusable components and guided by certain principles. To speed-up the design and development process, it needs to be Balanced, Consistent and Scalable. Design system is not a static document but rather an evolving system that supports the new, increasingly modular approach to digital product development.

**5.1 Data Flow Diagram**

A Data Flow Diagram (DFD) is a graphical representation of the flow of data through an information system, modeling its process aspects. A DFD is often used as a preliminary step to create an overview of the system without going into great detail, which can later be elaborated.

**5.1.1 Level 0 Data Flow Diagram**

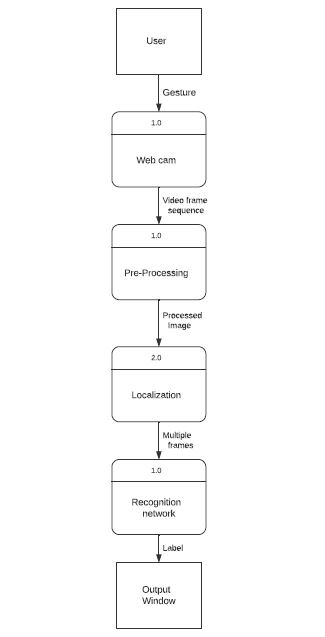
A level 0 data flow diagram (DFD), also known as a context diagram, shows a data system as a whole and emphasizes the way it interacts with external entities. Fig 5.1 shows the Level 0 Data Flow Diagram of the sign language recognition.

****

**Fig 5.1:** DFD Level 0 Sign Language Recognition

**5.1.2 Level 1 Data Flow Diagram**

A level 1 data flow diagram (DFD) is more detailed than a level 0 DFD. Fig 5.2 shows the Level 1 Data Flow Diagram of the sign language recognition.



**Fig 5.2:** DFD Level 1 Sign Language Recognition

**CHAPTER 6**

**TECHONOLOGY USED**

This system consists dynamic sign language recognition framework with deep neural network. Proposed architecture adopts deep convolutional neural networks with stacked temporal fusion layers as the feature extraction module and recurrent neural networks as the sequence learning module. Deep Neural Networks methods achieve high accuracy; however, they need a lot of training data. Creating Sign Language data can be time-consuming and costly. The motivation is to achieve comparable results with limited training data using deep learning for sign language recognition.

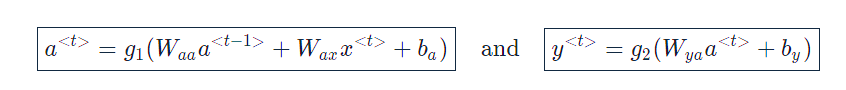
**6.1 CNN Model**

A Convolutional Neural Network (CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a Convolutional Neural Network is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, Convolutional Neural Networks have the ability to learn these filters/characteristics. The role of the Convolutional Neural Network is to reduce the images into a form which is easier to process, without losing features which are critical for getting a good prediction.

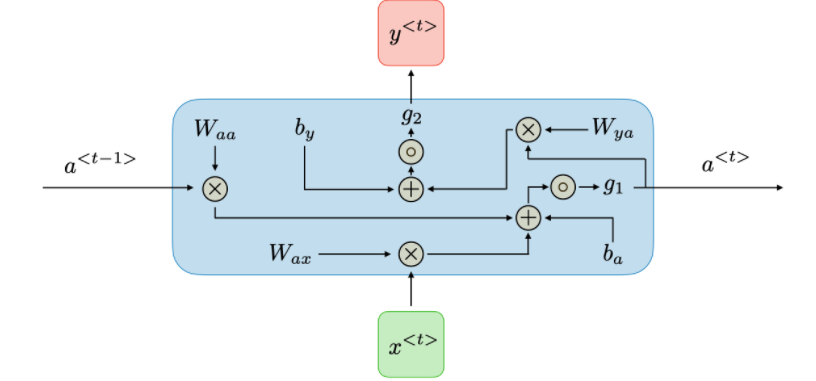
**6.2 RNN Model**

A Recurrent Neural Network (RNN) is a class of neural network, which allow previous outputs to be used as inputs while having hidden states.

For each time step t*­­*, the activation *a*<*t*> and the output *y*<*t*> are expressed as follows:



Where *Wax*​, *Waa*​, *Wya*​, *ba*​, *by*​ are coefficients that are shared temporally and  *g*1​, *g*2​ activation functions.



**Fig 6.1:** RNN repetitive block

The activation value (at) is the function of the activation value (at-1) from previous block and input value of present block (xt). RNN captures the sequential information present in the input data i.e. dependency between the words in the text while making predictions. RNNs share the parameters across different time steps. This is popularly known as Parameter Sharing. This Parameter Sharing is shown in Fig 6.1. This results in fewer parameters to train and decreases the computational cost.

**CHAPTER 7**

**CONCLUSION AND FUTURE SCOPE**

**7.1 Conclusion**

Hand gestures are powerful way for human communication, with lots of potential applications in the area of human computer interaction. Vision based hand gesture recognition techniques have many proven advantages compared with traditional devices. However, hand gesture recognition is a difficult problem and the current work is only a small contribution towards achieving the results needed in the field of sign language gesture recognition. This report presented a vision based system able to interpret hand gestures from the Indian Sign Language (ISA).

Videos are difficult to classify because they contain both the temporal as well as the spatial features. In the system there are two different models to classify on the spatial and temporal features. CNN was used to classify on the spatial features whereas RNN was used to classify on the temporal features. It is obtained that both the model combining gives accurate results. This shows that CNN along with RNN can be successfully used to learn spatial and temporal features and classify Sign Language Gestures. These two neural networks combined showed good result in terms of Recognition and Accuracy. This project achieved the goal for single gesture for single word and also achieved close to natural way of communication.

**7.2 Future Scope**

This work can be further extended in recognizing continuous sign language gestures with better accuracy. This method for individual gestures can also be extended for sentence level sign language. Also the current process uses two different models, training CNN followed by training RNN. Future work can be focused on combining the two models into a single hybrid model.

**References**

[1] P. V. V. Kishore, D. A. Kumar, A. S. C. S. Sastry and E. K. Kumar, "Motionlets Matching With Adaptive Kernels for 3-D Indian Sign Language Recognition," in IEEE Sensors Journal, vol. 18, no. 8, pp. 3327-3337, 15 April15, 2018, doi: 10.1109/JSEN.2018.2810449.

[2] H. Muthu Mariappan and V. Gomathi, "Real-Time Recognition of Indian Sign Language," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1-6, doi: 10.1109/ICCIDS.2019.8862125.

[3] T. Oliveira, N. Escudeiro, P. Escudeiro, E. Rocha and F. M. Barbosa, "The VirtualSign Channel for the Communication Between Deaf and Hearing Users," in IEEE Revista Iberoamericana de Tecnologias del Aprendizaje, vol. 14, no. 4, pp. 188-195, Nov. 2019, doi: 10.1109/RITA.2019.2952270.