A PROPOSED DESIGN AND IMPLEMENTATION OF SIGN LANGUAGE DETECTOR

**A PROJECT REPORT**

###### ***Submitted by***

**Pratistha Singh (21BAI10135)**

**Raghav Soni (21BAI10140)**

**Sarwagya Agarwal (21BAI10018)**

**Sumeet Deshpande (21BAI10172)**

**Tarushi Arora (21BAI10134)**

*in partial fulfillment for the award of the degree*

*of*

**BACHELOR OF TECHNOLOGY**

*in*

# COMPUTER SCIENCE AND ENGINEERING

****

**SCHOOL OF COMPUTING SCIENCE AND ENGINEERING**

**VIT BHOPAL UNIVERSITY**

**KOTHRIKALAN, SEHORE**

**MADHYA PRADESH – 466114**

OCTOBER 2022

**VIT BHOPAL UNIVERSITY, KOTHRIKALAN, SEHORE**

**MADHYA PRADESH – 466114**

**BONAFIDE CERTIFICATE**

Certified that this project report titled **SIGN LANGUAGE DETECTOR** is the bonafide work of **Pratistha Singh(21BAI10135), Raghav Soni(21BAI10140), Sarwagya Agarwal(21BAI10018), Sumeet Deshpande(21BAI10172), Tarushi Arora(21BAI10134)** who carried out the project work under my supervision. Certified further that to the best of my knowledge the work reported at this time does not form part of any other project/research work based on which a degree or award was conferred on an earlier occasion on this or any other candidate.

**PROGRAM CHAIR PROJECT GUIDE**

Dr. Suthir Sriram Dr. Lakshmi D

SCSE, VIT – Bhopal. SCSE, VIT – Bhopal.

The Project Exhibition I Examination is held on **03rd October 2022.**

**ACKNOWLEDGEMENT**

First and foremost, I would like to thank the Lord Almighty for His presence and immense blessings throughout the project work.

I wish to express my heartfelt gratitude to Dr Suthir Sriram, Head of the Department, School of Computer Science and Engineering for much of his valuable support encouragement in carrying out this work.

I would like to thank my internal guide Dr. Lakshmi D, for continually guiding and actively participating in my project, giving valuable suggestions to complete the project work.

I would like to thank all the technical and teaching staff of the School of Computer Science and Engineering, who extended directly or indirectly all support.

Last, but not least, I am deeply indebted to my parents who have been the greatest support while I worked day and night for the project to make it a success.

**LIST OF FIGURES AND GRAPHS**

|  |  |  |
| --- | --- | --- |
| **FIGURE NO.** | **TITLE** | **PAGE NO.** |
| 1. | American Sign Language | 10 |
| 2. | 21 Hand Landmarks | 21 |
| 3. | Software Architectural Diagram | 22 |
| 4. | User Interface Design | 22 |
| 5.1 | Code Stub 1 (Modules and Input) | 25 |
| 5.2 | Code Stub 2 (Hand Tracking) | 26 |
| 5.3 | Code Stub 3 (Conditions) | 27 |
| 5.4 | Code Stub 4 (Output) | 27 |
| 5.5 | Code Stub 5 (Extra Functionalities) | 28 |
| 6. | Testing | 31 |
| 7. | Word Formation | 31 |

**ABSTRACT**

**A proposed design and implementation of Sign Language Detector**

**Pratistha Singh (21BAI10135), Raghav Soni (21BAI10140), Sarwagya Agarwal (21BAI10018), Sumeet Deshpande (21BAI10172), Tarushi Arora (21BAI10134)**

**School of Computer Science and Engineering**

**Vellore Institute of Technology – Bhopal.**

There are more than 65 million people suffering from Significant Auditory Impairment in the world, and out of these approximately 60 million of them are in India. As such, there is a need of an interface that converts the Sign Language to normal English Alphabets to remove the communication gap. Our **design** (code) tries to solve this problem with a whole unique approach giving an accuracy of 90 - 95%.

The design implements Hand-Tracking Module of cvzone which gives us the orientation of the hand and 21 hand landmarks using which we provide the output. The input is the real time video of the user and the output is the letter of the English alphabet that the user shows. The code then checks for certain conditions using the 21 hand landmarks and gives the output.

This method is both, unique and innovative. Although prone to image noises and lightning conditions, our design was able to detect the gestures of the users with an accuracy of 90 – 95%.

**CHAPTER-1:**

**PROJECT DESCRIPTION AND OUTLINE**

* 1. **Introduction**

Communication is very crucial to human beings, as it enables us to express ourselves. We communicate through speech, gestures, body language, reading, and writing or through visual aids, speech being one of the commonly used among them. However, unfortunately, for the speaking and hearing-impaired people, there is a communication gap. Communication amongst humans is the root cause which binds them together. Communication is possible in various ways but one such way that has brought a major change to the world is the SIGN LANGUAGE.

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), Australian Sign Language (AUSLAN) and British Sign Language (BSL). Sign language is commonly used as the sole form of communication for people who are deaf or hard of hearing. It is basically the native language for deaf community which provides them full access of communicating within themselves and with others. Two forms of sign are used for communication in sign language: manual and non-manual. Manual signs involve fingers, hands, arms, and non-manual signs involve face, head, eyes, and body. Sign Language consists of fingerspelling, which spells out words character by character, and word level association which involves hand gestures that convey the word meaning.

* 1. **Motivation for the work**

There are more than 65 million people suffering from Significant Auditory Impairment in the world, and out of these approximately 60 million of them are in India. As such, there is a need to develop such a system that can translate the Sign Language and thus can help the hearing-impaired community to communicate better.

The main aim of our project is further motivated by the isolation that is felt within the deaf community. Loneliness and depression exist in higher rates among the deaf population, especially when they are immersed in a hearing world.

Being able to recognize Sign Language is an interesting computer vision problem while simultaneously being extremely useful for deaf people to interact with people who don’t know how to understand American Sign Language (ASL)**.**

* 1. **Problem statement**

Speech impaired people use hand signs and gestures to communicate which others face difficult to understand. Hence there is a need of a system which recognizes the different signs, gestures, and conveys the accurate information. It bridges the gap between physically challenged people and the non-disabled people.

* 1. **Objective of the work**

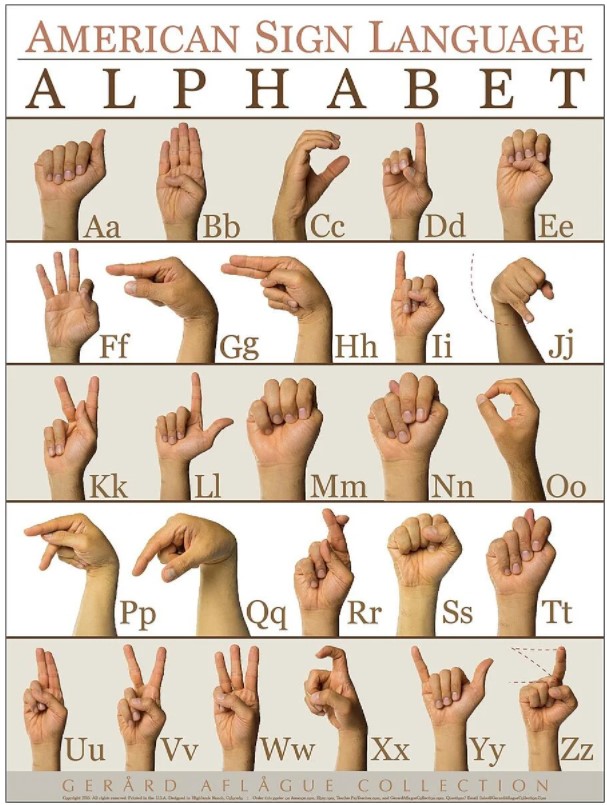
A functioning signing recognition system could provide a chance for the inattentive communicate with non-signing people without the necessity for an interpreter. During this project our aim is to develop a system which may classify signing accurately.

The objective of this project is to build an interface which will help to classify which letter of the American Sign Language (ASL) is being signed, given an image of a signing hand. Such a translator would greatly lower the barrier for many deaf and mute individuals to be able to better communicate with others in day-to-day interactions.

* 1. **Summary**

Difficulties are faced by the hearing-impaired community every now and then, and when immersed into a hearing world, their difficulties are further incremented.

Sign language is the only tool of communication for such person. Sign Language is a boon for the physically challenged people to express their thoughts and emotions. As the speech impaired and deaf people need a proper channel to communicate with others. There is a need of a prototype. Our project, hence is aimed at converting Sign Language gestures into a readable form.



**Fig. 1. – American Sign Language**

**CHAPTER-2:**

**RELATED WORK INVESTIGATION**

**2.1 Introduction**

People who are deaf or partially deaf have trouble expressing themselves and interacting with others. The main form of communication for the deaf and mute community is sign language. There is no universal sign language used by deaf people worldwide. However, unlike spoken languages, which vary from place to region, sign languages are not universal. When speaking or communicating, these individuals must utilize hand gestures or sign language. We then looked at some prior research that is relevant to our study. For example, there are two fundamental approaches used in sign language recognition: glove/device based, and vision based.

**2.2 Core Area of the Project**

The primary objective of our project is divided into two parts:

1. In the first part of the project, we have used HandTracking where computer vision has been used to detect a hand from an input image from the user. This allows us to show various gestures after detecting the hand movement.
2. In the second part we have used a method of determining the hand coordinates for each specific point of hand which thus provides us with an accurate alphabet when a hand gesture is shown.

**2.3 Existing Approaches/Methods**

2.3.1 Approaches /Methods -1: Glove Based Model

All procedures in this project were carried out by referring to the American Sign Language (ASL) Portal. Fingerspelling is mostly employed in the ASL manual alphabet to spell out names or English terms that do not have an established sign. There are 26 ASL alphabets in the database, with the rest being static gestures. The performer's hand motions are captured using a low-cost hand glove circuit constructed with numerous accelerometers. It generates each finger's flexion, the hand's movement and orientation, and the electrical output from the hand's muscle actions. The system recognizes gestures live, which means that the system takes the real-time signal from the gloves as an input and returns the matched gesture. It is entirely reliant on data.

2.3.2 Approaches /Methods -2: Vision Based Model

The Sign Language Recognition Prototype is a real-time vision-based system with the goal of recognizing the Sign Language alphabet. The prototype's goal was to verify the validity of a vision-based system for sign language recognition while also testing and selecting hand features that could be employed with machine learning algorithms for usage in real-time sign language recognition systems. To do so, the user must stand in front of the camera and make sign language movements, which the system will read and classify on the right side of the interface. Only one camera, a Kinect camera, is used in the implemented solution.

2.3.3 Approaches /Methods -3: Kinect Based Model

The system uses a Dynamic Time Warping (DTW) algorithm to recognize specific gestures and an off-the-shelf software application to generate audible language. Microsoft® Kinect is the primary instrument for capturing a user's video stream. With a 91 percent accuracy, the suggested technique can successfully detect gestures stored in the lexicon. Custom gestures can be defined and added to the proposed system. According to an experiment in which ten persons with disabilities used the system to communicate with five others who did not have a handicap, 87 percent believed that it was useful.

**2.4 Pros/Cons of stated Approached/Methods**

Glove based Approach

Glove based system supports a wider range of signs and modes. Also, this method uses a wireless transceiver system which was developed using a Gesture Vocalizer.

Vision-based Sign Detection

The Vision-based gesture extraction method is accurate for face and body gestures; however**,** it has a few issues. It frequently involves visual noises from different sources, including light, camera, color balancing, and the background. Numerous previous studies have shown that the error filter's dynamic environment prevented it from reconstructing the damaged regions. Additionally, this technique needs a real-time vision system in addition to a massive amount of processing.

Kinect-Based (Microsoft Kinect XBOX 360TM)

Kinect based method was developed initially for gaming purposes. But later Microsoft Kinect became widely recognized due to numerous industrialapplications in the fields of Computer Vision such as Gesture and Action recognition, VR (Virtual Reality), and Robotics. Although, this method is less widely used in industry due to its delayed responses whenthe feature size of image is large.Otherwise,it is quite an accurate method.

**2.5 Issues/Observations from investigation**

In comparison to sign language identification using our prototype, the glove-based, and kinect-based approaches either offer less accuracy or are more expensive and time-consuming. Furthermore, the initial architecture of our prototype offered a 90% accuracy. The glove-based strategy, on the other hand, only offered 75% accuracy. Additionally, our prototype gesture detection is very less expensive than Microsoft Kinect approaches.

**2.6 Summary**

No method is perfect, and our prototype is no exception, but the uniqueness of it makes it to stand out, thus giving an edge. There are methods which can provide more efficiency than this prototype but then a lot of effort, time and power is used to implement such methods.

**CHAPTER-3:**

**REQUIREMENT ARTIFACTS**

**3.1 Introduction**

Object Detection is a computer technology that deals with image processing and computer vision, it detects and identifies objects of various types such as humans, animals, fruits & vegetables, vehicles, buildings etc. Every object in existence has its own unique characteristics which make them unique and different from other objects and so does our hands. The basic requirements are environmental requirements and technical requirements.

**3.2 Hardware and Software Requirements**

3.2.1 Hardware Requirements

* PC or Laptop (Mac or Linux or Windows)
* A good quality camera (captures color image with minimum 480x640 resolution)

3.2.2 Software Requirements

* A Python IDE (preferably PyCharm) to run python code or any other text editors that can run python code (we have used VS Code)
* Installation of all the dependencies such as important libraries, etc.

**3.3 Specific Project Requirements**

3.3.1 Environmental Requirements

* A well-lit room is best suited for this project.
* The background of the user should be as free from other objects as possible to provide more accurate results. (a wall or something is best.)
* No other person should be in the frame of the video as it may affect the performance of the code.
* The user should show the gesture by keeping their hands at 30 to 60cms from the camera.

3.3.2 Technical Requirements

* Important libraries should be installed – opencv/cv2, mediapipe, cvzone, numpy, PIL, tkinter, etc.
* A laptop with minimum 4GB of Ram.

**3.4 Summary**

The above-mentioned requirements are not the only necessary and sufficient conditions but if fulfilled, the code can recognize the gestures with the mentioned accuracy i.e., 90 – 95%.

**CHAPTER-4:**

**DESIGN METHODOLOGY AND ITS NOVELTY**

**4.1 Methodology and Goal**

The entire project is divided into two parts: HandTracking and Feature Matching.

1. HandTracking: Hand tracking is the process in which a computer uses computer vision to detect a hand from an input image and keeps focus on the hand’s movement and orientation. Hand tracking allows us to develop numerous programs that use hand movement and orientation as their input.
2. Feature Matching: HandTracking gives us the way to mark 21 hand landmarks (features) on our hands with the help of which we detect gestures based on some conditions.

**4.2 Functional Module Design and Analysis**

Our design is based on mainly two modules – cv2 and cvzone. Both of these modules deal with Computer Vision (CV) and helps us to efficiently recognize the gestures.

* Cv2

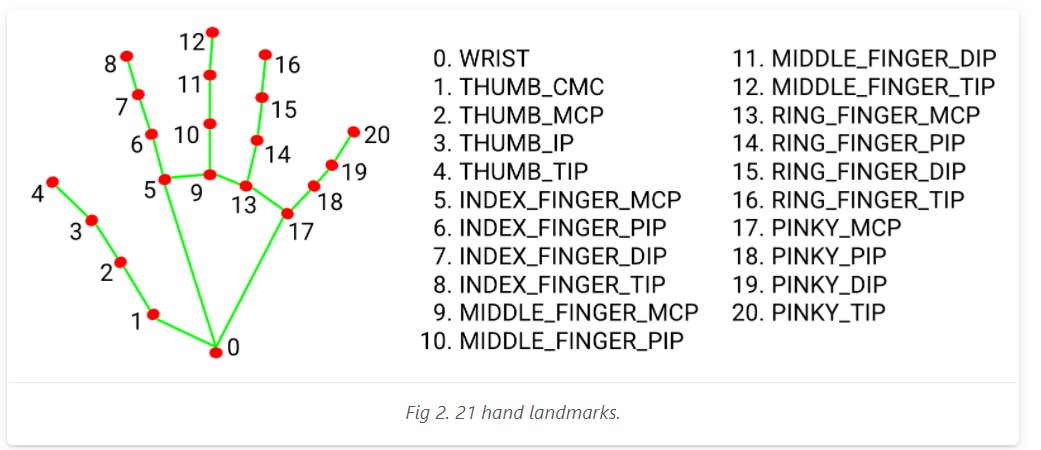
Cv2 is the module import name for opencv-python. OpenCV is a great tool for image processing and performing computer vision tasks.

We have used cv2 to capture video from our cameras. It is an open-source library that can be used to perform tasks like, writing things on the window/screen, making boxes, etc.

* Cvzone

Cvzone is Computer Vision package that makes its easy to run Image Processing and AI Functions. At the core it uses OpenCV and Mediapipe libraries.

MediaPipe is a Framework for building machine learning pipelines for processing time-series data like video, audio, etc. Using this, we can detect and track our hands. Cvzone gives us the ability to know the orientation and position of each joint in the fingers by providing us 21 hand landmark coordinates. It is with the help of these coordinates that we can achieve such a high accuracy.



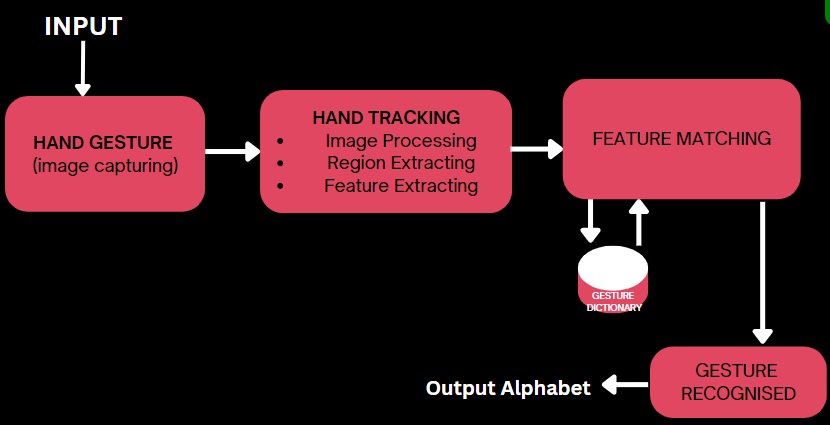
**Fig. 2. – 21 Hand Landmarks**

**4.3 Software Architectural Diagram**

Our prototype uses the 21 hand landmarks to recognize the shown gesture. This was made possible by a very simple but new and innovative approach. It was seen that every gesture in the American Sign Language (ASL) has a unique positioning and orientation of fingers and the hands. So, this led to the construction of conditions based on the signs of ASL (gesture dictionary), which was then used as a reference to detect and recognize the shown gesture.

This greatly increased the efficiency as well as the cost. The required time and effort reduced significantly. The computation power required was also lowered to bare minimum. The architectural diagram of the mentioned process

**Fig. 3. – Software Architectural Diagram**



**4.4 User Interface Design**

To make the design easy to use for the users, we made a small GUI using tkinter package of python which provides us with its vast libraries which helps us to create small GUI applications. The user design is simple yet elegant in its own way.

 The ‘Read Instructions’ button provides the user with all the important points that the user must keep in mind while using this prototype. ‘Start Application’ button starts the camera and hence the main code where it detects the gestures. ‘Show Gestures’ enables the user to see the gestures in the ASL which our prototype can recognize.

**Fig. 4. – User Interface Design**

**4.5 Summary**

The goal of this design/prototype is to detect Sign Languages that are present in the ASL. The method is unique as a whole and no one has ever tried it with this approach. The accuracy after all the testing comes out to be 90 – 95%.

**CHAPTER-5:**

**TECHNICAL IMPLEMENTATION & ANALYSIS**

**5.1 Outline**

The design is made possible using python programming language and its packages and libraries. The code is simple yet effective to recognize gestures and even form words out of it.

**5.2 Technical Coding and code solutions**

Text

Description automatically generated

**Fig. 5.1 – Code Stub 1 (Modules and Input)**

Text

Description automatically generated

**Fig. 5.2 – Code Stub 2 (Hand Tracking)**

Text

Description automatically generated

**Fig. 5.3 – Code Stub 3 (Conditions)**

Text

Description automatically generated

**Fig. 5.4 – Code Stub 4 (Output)**

Text

Description automatically generated

**Fig. 5.5 – Code Stub 5 (Extra Functionalities)**

**5.3 Test and Validation**

After testing our Gesture Detector (GD) we got to a conclusion that our accuracy is up to 90-95% as this algorithm is successful for comprehending static gestures. Signs like “J” and “Z” requires dynamic movements which requires model training, but the accuracy is not good enough compared to GD.

To make this approach stronger as well as user friendly we also added GUI to our application which will help user to get and easy idea how this works. This GUI consists of a user’s reader wherein the guide instructs the user certain rules and some precautions to work with ASL. The user’s guide can also be opened after showing a particular hand gesture. This GUI also have the image of how a person can show gestures by looking at it simultaneously making it. This app consists of all the features which can help the deaf and dumb at an preliminary level to understand ASL and can communicate with others without any barrier.

It doesn’t end here. The newness of our app keeps growing better. What if this app can display words too using just our fingers?

Yes, this app even has this extraordinary feature of combining a series of gestures and displaying a word. This approach has never been used before and this multiple features within one window makes this app special and unique.

**5.4 Summary**

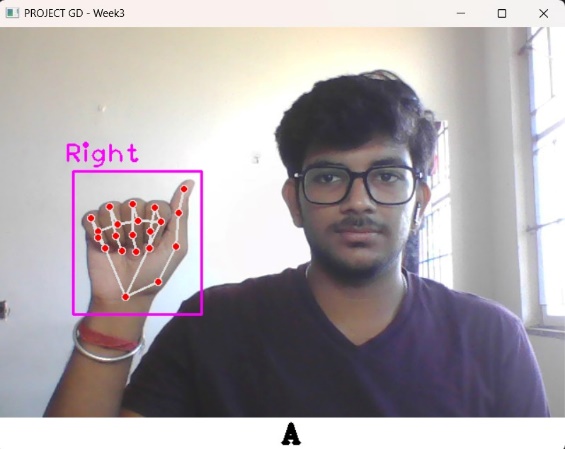
In short, this app is a one stop place for communication. This model is user friendly and easy to comprehend by user. This will surely bridge the gap of communication barrier to deaf and dumb community. The Validation is the process of exercising software with the intent of ensuring that the software system meets its requirements and user expectations and does not fail in an unacceptable manner.

**CHAPTER-6:**

**PROJECT OUTCOME AND APPLICABILITY**

**6.1 Outline**

Using tools learnt in computer vision and machine learning, we have developed a system that recognizes sign language gestures in real time. As a result of specifying conditions for the 21 3D landmarks, every condition can produce a unique letter of the English alphabet, and the following output was achieved (for alphabet A).



**Fig. 6. – Testing**

**6.2 Key Implementations of the System**

Except for dynamic gestures, our design recognizes gestures with an accuracy of 90-95%. An easy-to-use GUI makes it easy to operate the code. In addition to performing the same task, there are many other projects that can accomplish the same thing. What makes our project unique is that it offers the option of combining a series of gestures into a word.

Graphical user interface, application

Description automatically generated

**Fig. 7. – Word Formation**

**6.3 Project Applicability on real-world applications**

Our functioning signing recognition system provides a chance for the inattentive communicate with non-signing people without the necessity for an interpreter**.** It consists of all the features which can help the deaf and dumb at a preliminary level to understand ASL and can communicate with others without any barrier. In short this is a one stop place for communication.

**6.4 Inference**

Our goal with this project was to build an interface that would categorize which letters of the American Sign Language (ASL) alphabet were being signed given an image of a signing hand. The purpose of this project is to develop a possible sign language translator which is capable of translating communications from sign language to written and oral languages. In addition to improving communication with others on a daily basis, such a translator would greatly lower the barrier for many deaf and mute individuals

**CHAPTER-7:**

**CONCLUSIONS AND RECOMMENDATION**

**7.1 Outline**

Our Project GD facilitates communication processes for the hearing-impaired community by making them user-friendly and easily accessible. Specifically, we have implemented a method that makes use of the camera and Python libraries, as well as directly dealing with motion in the picture. To conduct our project, we examined the ASL signs carefully, extracted what they had in common, and separated them from each other.  
In addition to providing the user with a natural environment, this kind of vision-based solutions reduce the difficulties associated with rehabilitation for people with disabilities.

**7.2 Limitations/Constraints of the System**

Adequate lighting is required so that the Project GD can recognize the hand motions of the user. Our current system can detect only static gestures, limiting our ability to detect gestures in motion, so we're only able to detect 24 alphabets at a time, leaving 'J' and 'Z' undetected.  Furthermore, our code is still in the development stage, so to get accurate results, the user must show an exact hand gesture. Users should try to place their background as free of other objects as possible to provide more accurate results (a wall or something similar works well). Interference of any other person in the frame of video may affect the performance of our system.

**7.3 Future Enhancements**

Our approach could be extended by including non-alphabetic motions as well as all ASL alphabets in the gesture detection algorithm. The architecture of this project can also be extended to several other applications, such as hand gesture navigation control for robots. A sentence that is a more accurate translation of the language will be shown instead of letter labels. The letter detection system can be made more accurate by adding more training data. It is possible to develop this idea further to speak the signs. The project GD can also be developed as a web-based or app-based version, expanding the reach even further. In the future, we can enhance the recognition capability for various lightning conditions, which we encountered as a challenge in this project.

**7.4 Inference**

The objective of our project is to develop a real-time automatic system for recognizing sign language motions utilizing computer vision and machine learning techniques. We think our strategy is far more successful than training a model and inputting a huge number of photos merely to recognize characters.

**REFERENCES**

Base Paper:

[1] <http://cs231n.stanford.edu/reports/2016/pdfs/214_Report.pdf>

[2] <http://cse.anits.edu.in/projects/projects1920C6.pdf>

[3] <http://vision.stanford.edu/teaching/cs231a_autumn1213_internal/project/final/writeup/distributable/Chen_Paper.pdf>

Other References:

[1] Senior’s Base paper

[2] <https://google.github.io/mediapipe/solutions/hands.html>