**Hand Gesture Recognition (ASL Detection)**

**Department of Computer Science**

**Section: B**

**Semester: 6th**

SEMESTER PROJECT PROPOSAL

**By**

**Ayesha Shahzad (02-134201-083)**

**Sadia Urooj (02-134201-081)**

**Syeda Aqsa Ashraf (02-134201-076)**

**Proposed Project Supervisor(s):**

Miss Salas Akbar

Miss Soomal Fatima

**Department of Computer Science**

**Bahria University Karachi Campus**

# ABSTRACT

Computer vision is one of the most human assistive technologies Ai has blessed the humanity with. Hand gesture recognition with computer vision has a lot of practical implementations in the society such as helping visually impaired people or medical imaging. One of its uses can be implemented on Sign Language detection. Sign language is a mean of communication for the people who are deaf and dumb. Any movement of body or particularly hands and arms are used to relay a message in sign language. Different countries use different kinds of sign language, these languages are formulated according their own grammar, sentence structures and alphabet. Although this is a widely used method by most differently abled people but many individuals still don’t understand or learn this language when it’s not of any use for them. In order to help people who have had a hard time in maintaining basic communication in the society we use computer vision to translate these languages so that the people who do not learn it can understand the other person without any hassle. In this project, we will implement hand gesture recognition to detect American sign language with Convolutional Neural Networks and OpenCV. This is a real time method in which a sign language user shows gestures in a camera and the ML trained model acts as an interpreter and translates it to English language. We gather some data of hand gestures showing sign language and train the model with convolutional filters and predict the outcomes.

Sign language detection in real time can assist people with hearing impairments in a number of manners such as real time translation in a zoom meeting or conference. Google translate is also working on different sign language translations. As we all know speech recognition systems currently are very advanced and accurate, they’re being widely used in homes, offices and personal devices such as ALEXA, GOOGLE ASSISTANT, SIRI. However, sign language recognizers are far from the required accuracy and usability. A number of machine learning algorithm such as convolutional filters are tested and trained to achieve a well maintained and accurate model in recent years to help the Deaf and Dumb community build a bridge towards the majority hearing community. By dissecting images into pixels with labels or tags, a CNN aids a machine learning or deep learning model to recognize. It creates predictions about what it is looking at by performing convolutions on the labels, which is a mathematical operation on two functions to create a third function. Until the predictions start to come true, the neural network conducts convolutions and evaluates the accuracy of its predictions repeatedly. Then, it recognizes images similar to how a human does. In this project, we will acquire a dataset consisting of different hand gestures representing a letter in ASL the data acquired will be preprocessed and features will be extracted, this data will then be trained with convolutional neural network algorithm (such as SqueezNet) to achieve maximum accuracy specially in the gestures that are similar to each other.

**Content**

|  |  |
| --- | --- |
| CHAPTER 1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. . . . .  Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. . . . .  1.1 Problem Area . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . ..  1.2 Scope . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .  1.3 Expected Time . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .  1.4 Specific Aims & Deliverables . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .  1.5 Tools and Benefits . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .  1.6 Description . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .  1.7 Motivation . . . . . . . . . . . . . . . . . . . . . . . . . . . ­­. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | -1-  -1-  -1-  -1-  -1-  -1-  -2-  -2-  -3- |
| CHAPTER 2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .  Literature Review . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .  2.1 Description of related work . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | -3-  -3-  -3, 4- |
| CHAPTER 3 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .  Problem Definition . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .  3.1 Sign Language Recognition with computer vision . . . . . . . . . . . . . . . . . . . . . . . . .  3.2 Activity Plan . . . . . . . . . . . . . . . . . . . . . .. . . . . . . . . . . . . . . .. . . . . . . . . . . . . . . . | -4-  -4-  -4-5-  -6- |
| REFERENCES . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | -6- |

**CHAPTER 1**

## Introduction

## 1.1 Problem Area

## There are many applications where hand gestures are used for interaction with the system like videogames and medical equipment etc. These hand gestures can also be used by handicapped people to interact with the systems. Classical interaction tools like keyboard, mouse and touchscreen, etc., may limit the way we use the system. Gestures can interpret the same functionality without physically interacting with the interfacing devices. The problem lies in the understanding of these gestures by machine as for different people, the same gestures may look different. This problem can be resolved by the use of Deep learning approaches. Convolution neural networks (CNN) are providing to be the ultimate tool to process the image data for these systems.

## 1.2 Scope

Hand gesture recognition is one of the active research areas in the field of human-computer interface due to its flexibility and user friendliness. The gesture recognition technique is used to develop a system that can be used to convey information among disabled people or for controlling a device. In this project we will be building a model on sign language detection that takes the dataset in real time from the camera and then apply the required preprocessing techniques to filter out all the unnecessary information from the image data after that the pre-processed data is passed through a classifier which predicts the class of the hand gestures.

* **Project Objectives:**

The main motives of our project are as follows:

* + To build a model that can efficiently collect the data from the camera in real-time.
  + To make effective uses of CNN to training the required model for prediction.
  + To get approximate predictions from the trained model.

## 1.3 Expected Time

The project is expected to be completed in 2 months

## 1.4 Specific Aims and Deliverables

* To build a hand gesture recognizer that predicts an ASL gesture and translate it into text.
* To increase the accuracy of the model.

## 1.5 Tools and Benefits

**Tools:**

* Google Colab
* Tensor Flow
* OpenCV

**Benefits:**

1: Education and life improvement for children, elderly and deaf-mutes. Through some human-computer interfaces, can complete the natural communication between children, the elderly or deaf-mutes and the computers, and thus improve their ability of education. In addition, by using hand gesture recognition-based human-computer interfaces, a communication channel between the normal peoples and deaf-mutes can be established, so that normal peoples can “listen” and understand the “words” of deaf-mutes.

2: Application on smart home appliances and control field. Among the computer controlled means, hand is regarded as a flexible and efficient controlling way, application on hand gesture in the field of control has obtained some results, such as video cameras controlled by gesture command such as “zoom”, “panoramic” and “tilt”.

Jaguar company introduced a vehicle control system based on hand gesture recognition.



Figure 1: Conceptual map in response to incoming phone calls by driver hand gestures.

3: Demonstration and study of robot. By studying the mechanism of human visual language from the perspective of cognitive science to improve the human language understanding ability of computer.

## 1.6 Description

In this project, we will implement hand gesture recognition to detect American sign language with Convolutional Neural Networks and OpenCV, We will acquire a dataset consisting of different hand gestures representing a letter in ASL the data acquired will be preprocessed and features will be extracted, this data will then be trained with convolutional neural network algorithm (such as SqueezNet) to achieve maximum accuracy specially in the gestures that are similar to each other.

Classification of Images

Convolutional Filter Training

Data Acquisition (Vision Based image collection)



Outcome Prediction

Feature Extraction

Data Preprocessing

Figure 1.1: Process of the model training and outcome generation

## 1.7 Motivation

A desire for contactless sensing and hygiene concerns are the top drivers of demand for touchless technology. Gesture recognition can also provide better ergonomics for consumer devices. During the coronavirus pandemic, it’s not surprising that people are reluctant to use touchscreens in public places. Moreover, for drivers, tapping a screen can be dangerous, as it distracts them from the road. In other cases, tapping small icons or accidentally clicking on the wrong field increases frustration and makes people look for a better customer experience. All these problems which we figured out motivated us to work on hand gesture recognition. It is just the next step in technological evolution, and it’s ideally suited for today’s consumer landscape, so our main aim is to develop a system which can detect the gestures and these gestures are widely used for conveying the information or to control the devices.

**CHAPTER 2**

## Literature Review

## 2.1 Description of Related Work

*Zaki, M.M., Shaheen, S.I.: Sign language recognition using a combination of new vision based features. Pattern Recognition Letters 32(4), 572–577 (2011)* [[1](#_References:)] Worked on American Language detection is presented in which they have used a vocabulary set of 30 words. The system was appearance-based recognition with a hand tracking recognition to classify. They used hidden Markov Model. An error rate of 10.91% is achieved on the RWTH-BOSTON-50 database.

Cooper, H., Ong, E.J., Pugeault, N., Bowden, R.: Sign language recognition using sub-units. The Journal of Machine Learning Research 13(1), 2205–2231 (2012) [[5](#_References:)]. The approach in this paper uses the Microsoft Kinect to extract appearance-based hand features and track the position in 2D and 3D. The classification results are obtained by comparing a hidden Markov model (HMM) approach with sequential pattern boosting (SP-boosting). This resulted in an accuracy of 99.9% on 20 different isolated gestures on their specifically constructed data set and 85.1% on a more realistic one with 40 gestures.

**CHAPTER 3**

## Problem Definition

## 3.1 Hand Gesture Recognition.

Our Hand Gesture recognition model is implemented in four steps:

**Step: 1 - Data Collection:**

In this step, we collect images showing hand gesture which will represent each letter of the alphabet.

**Step: 2 - Preprocessing:**

After data collection the data set images that we obtain are in the colored form, and as colored images have a lot of features which are not required for our project because feeding colored images will take a lot of time and resolution for training the model also more computational power is required for processing the color images. So, in preprocessing step we do following steps:

* Converting the obtained images in (RGB) form to Black and White form in order to convert the 3 Channel (RGB) image to single channel.
* As we only require a sharp boundary of the hand image, for that we put thresholding (technique to detect the boundary of an object) on our black and white image with a Gaussian Blur.
* When we put neural network on preprocessed images it will provide more accurate results.

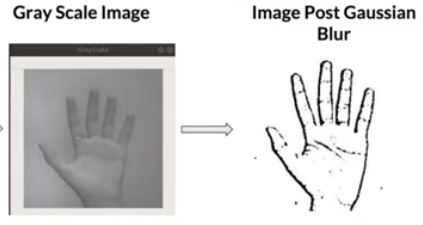


Figure 1: Conversion of Black and white image to only sharp edges image

**Step: 3 - Training:**

In this step we will implement the Convolutional neural network (CNN) model on the preprocessed data. In this state all the layers to be implemented on the data is defined according to the

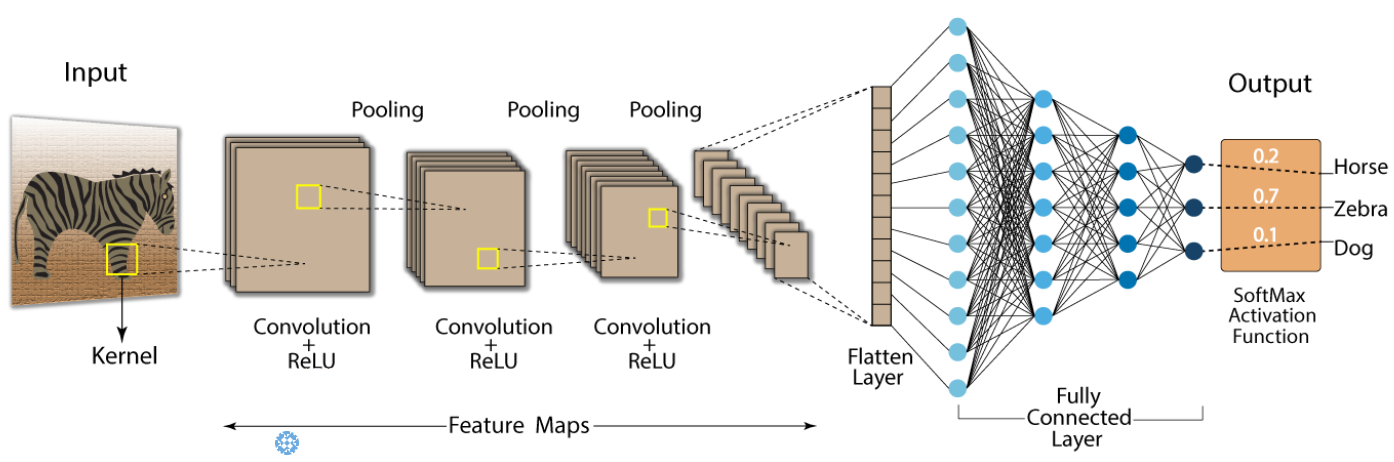


Figure 2: Convolutional neural network (CNN)

Tracking of people provides a significant

Advantage to identification since we can apply the

Principle of continuity of identity [2]. This says that, while

we may only be able to identify a person occasionally

(such as when we have a good view of their face, when

they swipe an ID badge, or when they speak into a

Telephone), if wecan reliablytrack the person, we know

That all identifications associated with the track relate to

The same person and applythroughout the track. Several

(Fallible) identification methods applied at different times

And places can thus be combined and corroborated.

Tracking of people provides a significant

Advantage to identification since we can apply the

Principle of continuity of identity [2]. This says that, while

we may only be able to identify a person occasionally

(such as when we have a good view of their face, when

Tracking of people provides a significant

Advantage to identification since we can apply the

Principle of continuity of identity [2]. This says that, while

we may only be able to identify a person occasionally

(such as when we have a good view of their face, when

**Step: 4 - Prediction:**

Using the trained model, we predict the hand gesture and predict the accuracy of the model.

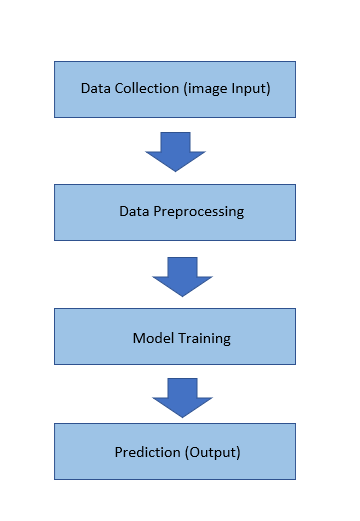
****

Figure : Project Flow Diagram

**3.2 Activity Plan**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **S. No.** | **Activity** | **June** | **July** | | **Aug** | **Sep** | **Oct** | **Nov** | **Dec** | **Jan** | **Feb** | | **March** | **April** | | | **May** |
| **1** | **Literature Review** |  |  | |  |  | **3 months** | | |  |  | |  |  | | |  |
| **2** | **Data Collection** |  |  |  | |  |  | **2 months** | |  |  | |  |  | | |  |
| **3** | **Preprocessing**  **Data** |  |  | |  |  |  |  | **2 months** | |  | |  |  | | |  |
| **4** | **Feature extraction** |  |  | |  |  |  |  | **2 months** | |  | |  |  | | |  |
| **5** | **Training the Model** |  |  | |  |  |  |  | **2 months** | |  | |  |  | | |  |
| **6** | **Final Testing of model** |  |  | |  |  |  |  | **2 months** | |  | |  |  | | |  |
| **7** | **Report writing** |  |  | |  |  |  |  |  | **1 month** |  |  | | |  |  | |

Tracking of people provides a significant

Advantage to identification since we can apply the

Principle of continuity of identity [2]. This says that, while

we may only be able to identify a person occasionally

(such as when we have a good view of their face, when

they swipe an ID badge, or when they speak into a

Telephone), if wecan reliablytrack the person, we know

That all identifications associated with the track relate to

The same person and applythroughout the track. Several

(Fallible) identification methods applied at different times

And places can thus be combined and corroborated.

# References:

[1] Zaki, M.M., Shaheen, S.I.: Sign language recognition using a combination of new vision-based features. Pattern Recognition Letters 32(4), 572–577 (2011)

[2] <https://gongster.medium.com/how-does-a-neural-network-work-intuitively-in-code-f51f7b2c1e3f>

[3] https://medium.com/nybles/a-brief-guide-to-convolutional-neural-network-cnn-642f47e88ed4

[5] Cooper, H., Ong, E.J., Pugeault, N., Bowden, R.: Sign language recognition using sub-units. The Journal of Machine Learning Research 13(1), 2205–2231 (2012)