Sign Language Recognition

Project Phase I (CSD 401)
Mid Semester Examination - 1 (September 27, 2021 , Monday)

PROJECT TEAM MEMBERS

1. Sanskruti Nakhale BT18CSE015
2. Niraj Agrawal BT18CSE035
3. Tanmay Ganvir BT18CSE036
4. Nisarg Gogate BT18CSE040
5. Ketan Sarode BT18CSE044

NAME OF SUPERVISOR

Dr. Umesh Deshpande

Contents

Introduction

The sign language is used widely by people who are deaf-dumb these are used as a medium for communication. A sign language is nothing but composed of various gestures formed by different shapes of hand, its movements, orientations as well as the facial expressions. There are around 466 million people worldwide with hearing loss and 34 million of these are children. ‘Deaf’people have very little or no hearing ability. They use sign language for communication. People use different sign languages in different parts of the world. Compared to spoken languages they are very less in number.

The goal of this project is to build a neural network, able to classify which letter of the American Sign Language(ASL) alphabet is being signed, given an image of a signing hand. This project is a first step towards building a possible sign language translator, which can take communications in sign language and translate them into written and oral language. 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.

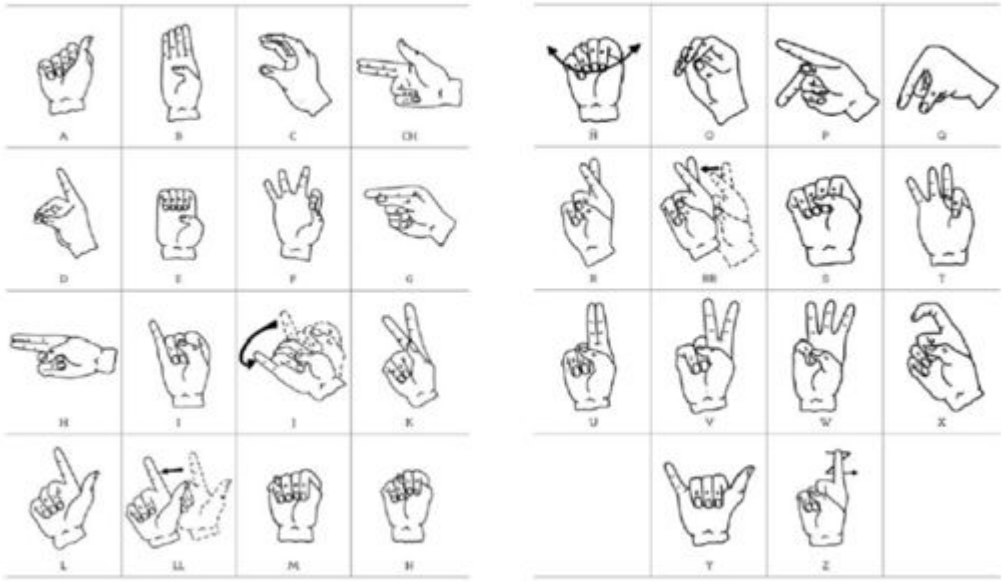
Deaf and dumb people make use of their hands to express different gestures to pitch their ideas with other people. Gestures are the nonverbally exchanged messages and these gestures are understood with vision. This nonverbal communication of deaf and dumb people is called sign language.

Introduction

Sign language is a visual language and consists of 3 major components:

Fingerspelling	Word level sign Vocabulary	Non-manual features
Used to spell words letter by letter.	Used for the majority of communication.	Facial Expressions and tongue, mouth and body position.

Pre-defined Hand Gestures:

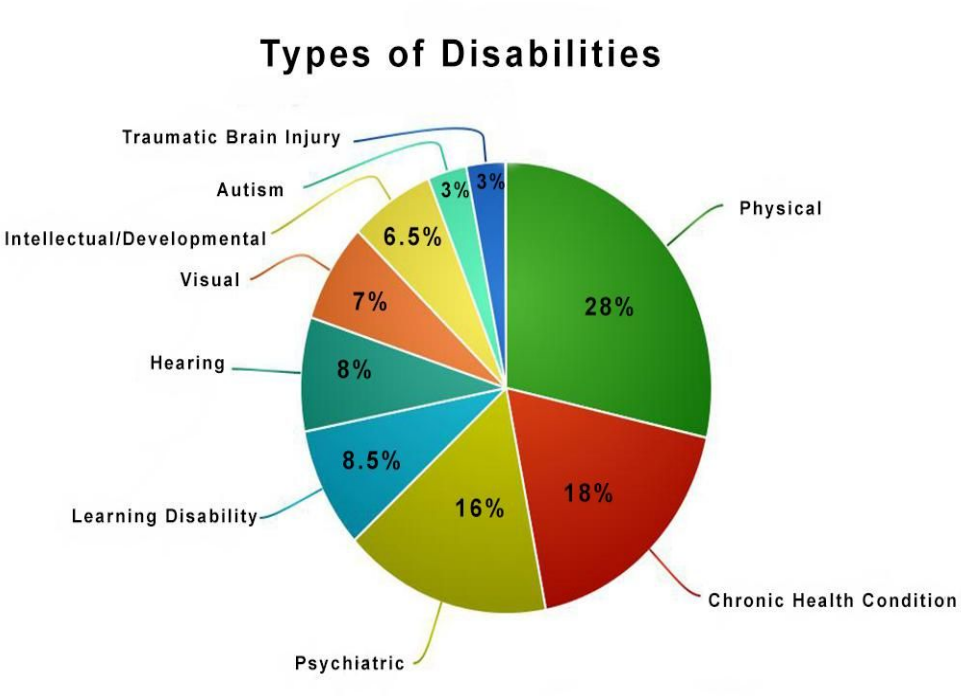


Panamanian Manual Alphabet (source: SENADIS, Lengua de Señas Panameñas)

Motivation

Communication is one of the basic requirement for survival in society. Deaf and dumb people communicate among themselves using sign language but normal people find it difficult to understand their language. Our project aims attaching the basic step in bridging the communication gap between normal people and deaf and dumb people using American sign language.

Effective extension of this project to words and common expressions may not only make the deaf and dumb people communicate faster and easier with outer world, but also provide a boosting Developing autonomous systems for understanding and aiding them.



Problem Statement

Given a hand gesture ,implementing such an application which detects pre-defined American sign language(ASL) in a real time through hand gestures and providing facility for the user to be able to store the result of the character detected in a txt file. It is a real time software system that will be able to recognize ASL hand-gestures using deep learning techniques.

This project aims to predict the 'alphanumeric' gesture of the ASL system.

Problem Formulation

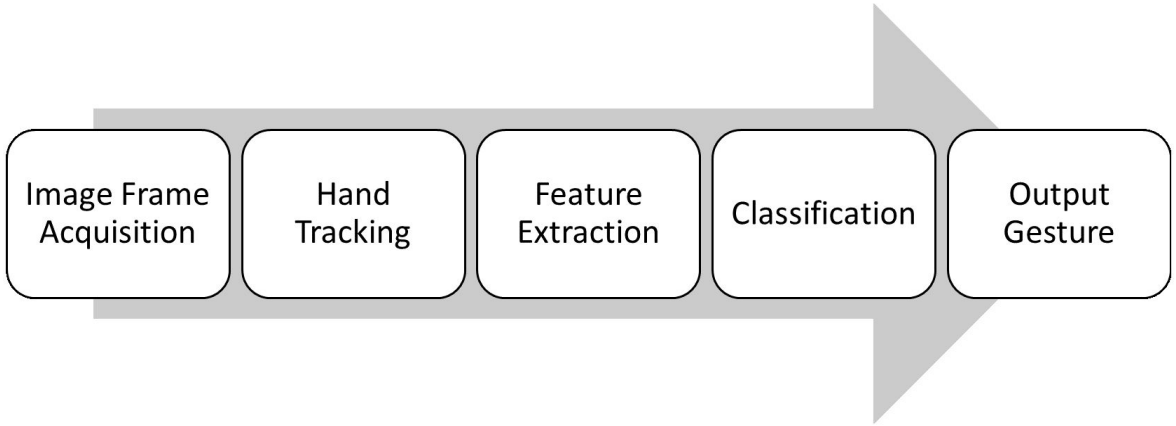
1. To generate large amount of appropriate dataset using vision based techniques like camera .
2. To apply appropriate image pre-processing techniques in order to remove the noise and obtain the ROI.
3. To design the model and architecture for Neural Network to train the pre-processed images and achieve the maximum possible accuracy.
4. To develop an algorithm to predict the gesture in real time using **tkinter** Python interface.

Literature Survey

In the recent years there has been tremendous research done on the hand gesture recognition.

With the help of literature survey done we realized the basic steps in hand gesture recognition are :-

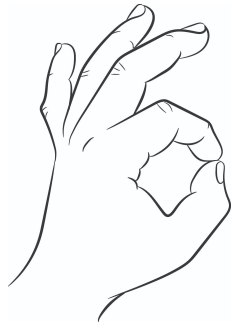
- Data acquisition
- Data preprocessing
- Feature extraction
- Gesture classification



Data Acquisition

Data Acquisition step consists of digitize the information from the user to make it usable to classification algorithms, the information from the user can be acquired in different ways:

- **Data Glove (Sensory Device)** : 5 sensor glove is commonly used in engineering VR(Virtual Reality). It uses electromechanical sensory devices to provide exact hand configuration and position. But they are expensive and not user friendly.
- **Camera (Vision Based):** Computer Vision is very popular but has a few drawback as well. There are a few factors which need to be taken into account.
    - Image Resolution - Better the camera more the pixels and hence more the data our model has to train.
    - Environmental Noise - Position of camera, light, background of image also plays a big role.



We plan to use Vision based Data Acquisition as it can be widely used because of low cost and considering high popularity of camera phones.

Data Preprocessing and Extraction

The data acquired by visual approach must be cleaned. We need to cope with the large variability of human hands appearance due to huge number of hand movements, **skin-colour** and **viewpoints**.

Popular used Morphical filters - Non Local Mean Filter (NLMF), Contra harmonic Mean Filter (CMF), Spatial Filter (SF) and Temporal Median Filter (TMF)

Morphological Filtering: After multiple steps of filtering hand will be represented by 1's and background by 0's. But this kind of filtering is more likely to not give proper results as in representation of A and M the hand gestures will look too similar.

A

M

Original RGB image

Skin color segmented image

Binary converted image

Morphologically dilated image

Morphologically filled hole image

Data Preprocessing and Extraction

Edge Detection and **Image Enhancement:**

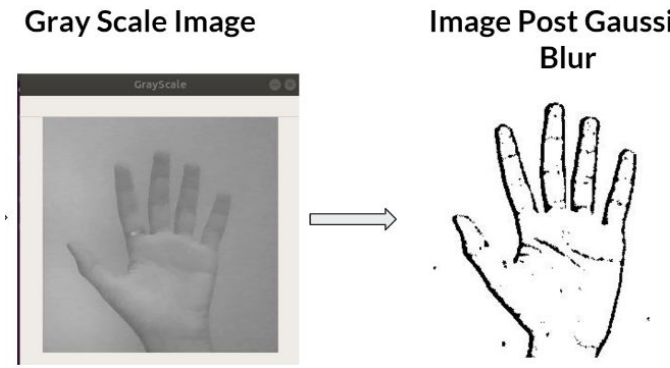
Adaboost detection - To differentiate face from hand based on colour.

Convert to Grayscale - RGB (0-255) to GrayScale(0-8)bits.

Gaussian Filtering - This can be used to reduce noise and helps in edge detection.

Edge detection are further highlighted using Prewitt operator, Sobel operator and Laplacian operator.

It is observed that colour segmentation results vary drastically with change in colour of background. Moreover we need to classify and train the model on similar looking gestures so rather than using random background we use a stable single color background.



Gesture Classification

To classify input hand gesture according to currently relevant hand gestures. To do this various methods are present like :-

- Using HMM(Hidden Markov Model)

  Used for classification of gestures. This deals with dynamic part of gestures. Gestures area unit collected through a sequence of video or picture streams. The goal is to acknowledge two categories of gestures one is deictic(gestures included **pointing, showing, giving, and reaching**, or some combination of these gestures) and other is symbolic(gestures used to make requests like putting index fingers on closed lips means to “be quiet”).

Gesture Classification

- Using CNN (Convolutional Neural Network):-

  They input the pre-processed image to a convolutional neural network model in order to train and predict the output. It takes the image's raw pixel data, trains the model, then extracts the features automatically for better classification.

According to research , when we have static gestures and large dataset, CNN proves to give high accuracy.

Our model gets pre-processed image as input.If a letter is detected for more than 50 frames* then the letter is printed and taken into consideration for forming the word. To detect blank or space just don’t show any gesture, so that when image is pre-processed model gets plain black image and on obtaining such image, it detects blank space and forms new string.

*A normal camera records 30fps. So even if the user is fluent in sign language he shall show the gesture for at least 1.5 seconds.

Work Done

- Dataset is chosen. We have taken an existing database present on web. Also we will capture ourselves showing different gestures.
- Decided preprocessing filters which are more likely to give a good edge detection for our model.
- On basis on different research papers, we chose to work with CNN model and decided the flow of project.

Work to be Done by the Next Evaluations

- Pre-process our self made dataset.
- Implement a CNN model using Deep Convolutional Networks.
- Training and Validation of model.
- Test the model.
- Peak validation accuracy of 85% minimum by changing filtration Techniques,Epoches,Methods.
- Make a GUI for on-time Gesture Detection using **tkinter** Python Interface.

Vision for the project

Vision is to make functional real time vision based American sign language recognition for Deaf and Dumb people using ASL alphabets.

We have to achieve considerable accuracy of at-least 85% on our dataset.

Then we try to improve our prediction after implementing the respective layers of algorithms in which we verify and predict symbols which are more similar to each other.

Planning to achieve higher accuracy even in case of complex backgrounds by trying out various background subtraction algorithms .

We are also thinking of improving the preprocessing to predict gestures in low light conditions with a higher accuracy.

References

[1]Manuel Eleazar Martínez-Gutiérrez, José Rafael Rojano-Cáceres, Edgard Benítez-Guerrero, Héctor Eduardo Sánchez-Barrera,"Data Acquisition Software for Sign Language Recognition" in Universidad Veracruzana, Facultad de Estadística e informática, Mexico

[2]Mathavan Suresh Anand, Nagarajan Mohan Kumar, Angappan Kumaresan, "An Efficient Framework for Indian SignLanguageRecognition Using Wavelet Transform" Circuits and Systems, Volume 7, pp 1874-1883, 2016

[3]Suharjitoa, Ricky Andersonb , Fanny Wiryanab , Meita Chandra Ariestab , Gede Putra Kusumaa ,"Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process-Output" in 2nd International Conference on Computer Science and Computational Intelligence 2017, ICCSCI 2017, 13-14 October 2017, Bali, Indonesia

[4]T. Yang, Y. Xu, and "A. , Hidden Markov Model for Gesture Recognition", CMU-RI-TR-94 10, Robotics Institute, Carnegie Mellon Univ.,Pittsburgh,PA, May 1994

References

[5]Pigou L., Dieleman S., Kindermans PJ., Schrauwen B. (2015) “Sign Language Recognition Using Convolutional Neural Networks”. In: Agapito L., Bronstein M., Rother C. (eds) Computer Vision - ECCV 2014 Workshops. ECCV 2014. Lecture Notes in Computer Science, vol 8925. Springer, Cham

[6]Byeongkeun Kang , Subarna Tripathi , Truong Q. Nguyen ”Real-time sign language fingerspelling recognition using convolutional neural networks from depth map” 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)

[7]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)

Thank You !!!