Sign Language Detection

Artificial Intelligence, MCSC 102
Department of Computer Science,
University of Delhi

Harsh Yadav
Purnima Kumar
Yashi Sharma

Overview

Problem Statement

We aim to develop a **computer- vision based AI system** that can accurately interpret and translate **sign language gestures** for the **English alphabet (A-Z) and numbers (0-9)** into corresponding **English text.** The system will empower individuals with hearing impairments to communicate effectively, bridging the gap between the deaf and hearing communities.

Background

The sign language is the fundamental communication method between people who suffer from hearing defects. Sign language can be considered as a collection of gestures, movements, postures, and facial expressions corresponding to letters and words in natural languages.

Static gesture

A static sign is determined by a certain configuration of the hand

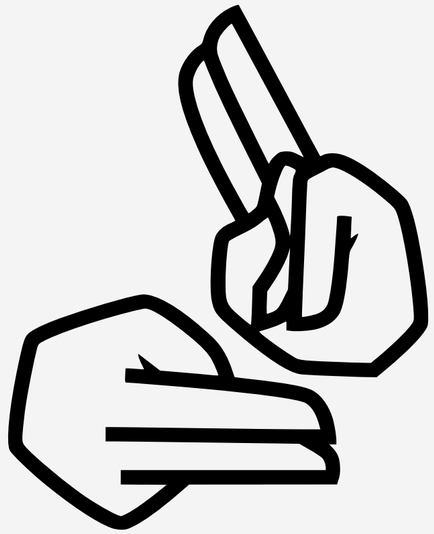
Dynamic gesture

Dynamic gesture is a moving gesture determined by a sequence of hand movements and configurations.

Related Work

Feature extraction, statistics and model
Template matching, feature extraction and analysis, active shape models, principal component analysis, linear fingertip models, casual analysis

Learning Algorithms
Neural network, Hidden Markov Model, Instance-based learning



Research on hand gestures

- **Electromechanical devices**
- **Machine vision & image processing**
  - **Visual based gestures with glove markers (VBGwGM)**
  - **Pure visual based gestures (PVBG)**

Data Set & Input/Output

01 About the data set

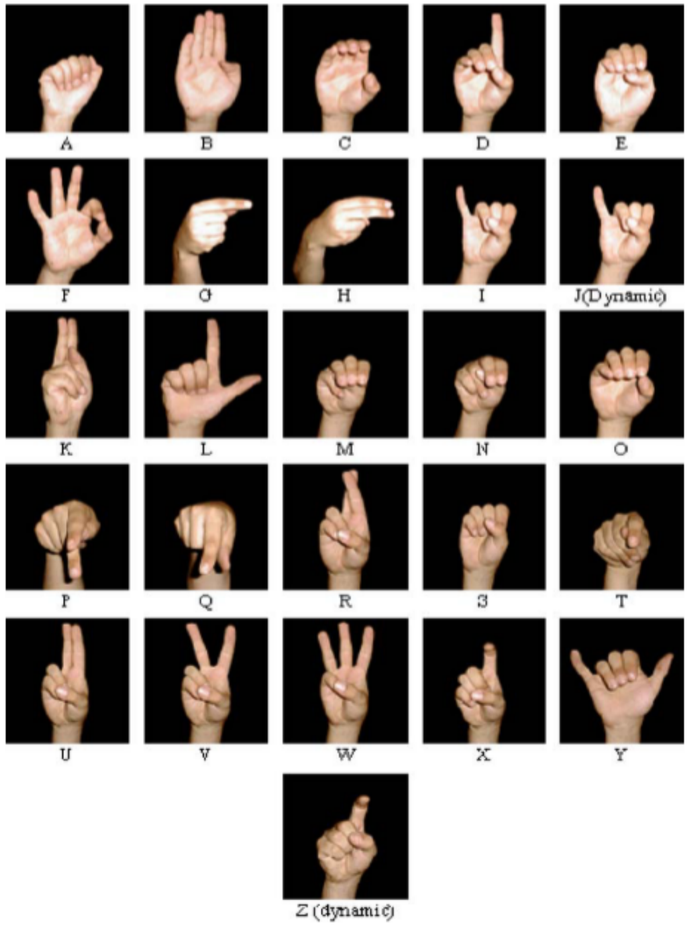
The data set contains 300 samples of hand sign images for American Sign Language with 15 images for each sign.

02 Input

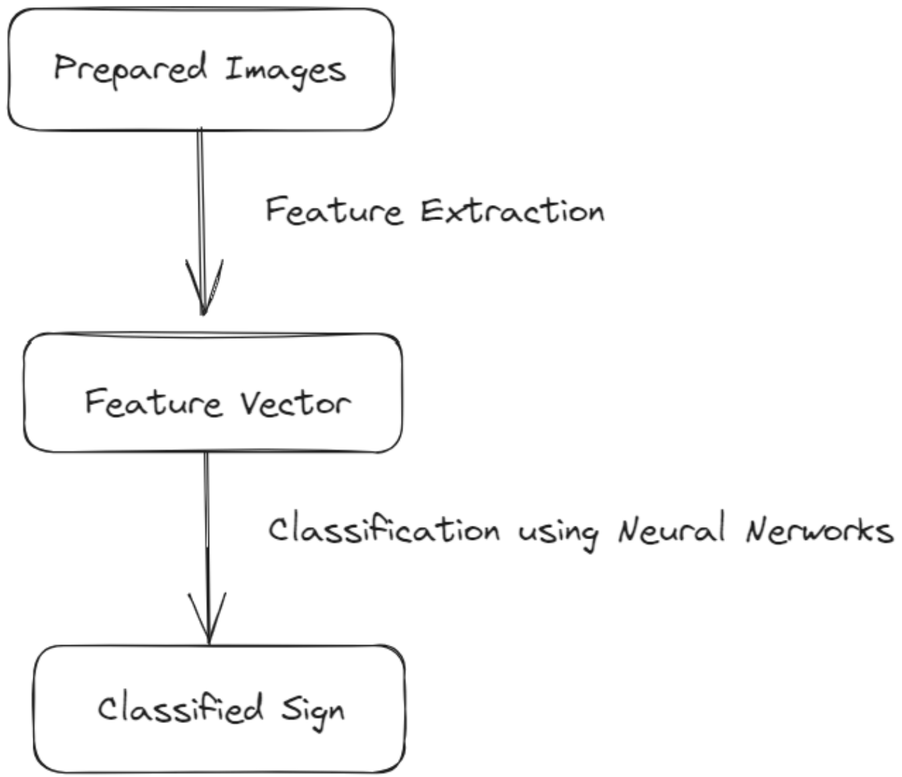
Images of **hand gestures** representing **American Sign Language (ASL)**.

03 Output

Accurate textual translations of English alphabet (A-Z) and numbers (0-9).



Work Flow



The work flow contains of two phases:

1. Feature Extraction
2. Classification

The Feature Extraction Phase:

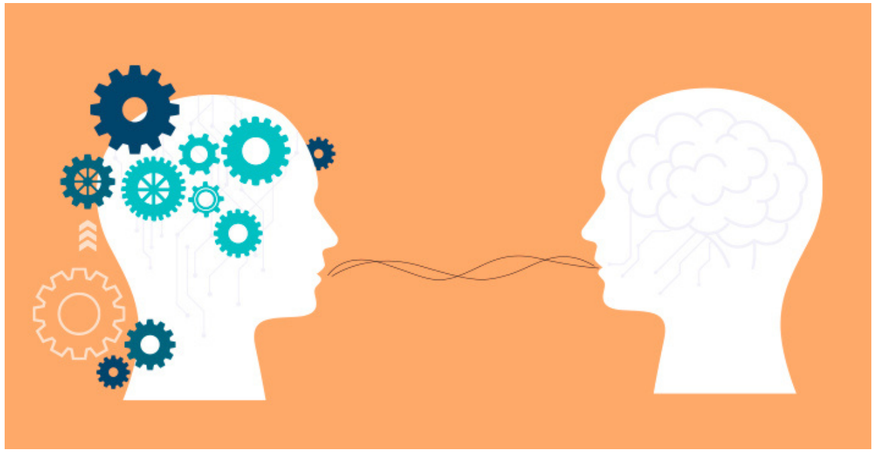
1. Resize the image
2. Convert RGB to gray scale
3. Edge detection using Canny Edge Detector
4. Hough Transform

The Classification Phase:

A neural network architecture is created and used to classify the images.

Transfer Learning Approach

- Transfer Learning is a process where the **weights of an existing pretrained model** that was created to perform a similar task are **retrained on a new data** in order to perform a new, but similar task.



- Used Microsoft COCO dataset, which is a large-scale object detection dataset containing **330,000 images** consisting of **80 object categories**.

1st Training

1

- **Faster_RCNN_Inception_V2**.
- Personal Computer
- Stops training after 200,000 iterations.
- Reached to 0.7 location loss.
- Took **127.8 hours** (5.3 days) in total.

Advantages

2

- **Correctly predicted** the sign.
- Successfully predicted the location of sign in image.

Fixing the Issue.

4

- Tried on **live video stream** from webcam.

Disadvantage

3

- **Mirroring.**

Result

5

- Mirroring problem solved.
- **Predicted correctly** most of the time.

Ran into another problem!

6

- The algorithm was **incredibly slow**.
- 1 frame/sec only.

2nd Training

1

- **SSD_Mobilenet_V2**
- Google Colab
- took **12 hours** of GPU time(10x faster).

Advantage

2

- Model was **significantly faster**.
- Could run a live video stream seamlessly.

Diadvantage

3

- **Not accurate** as previous one.

References

- American sign language (ASL) recognition based on Hough Transform and Neural Networks. Qutaishat Munib, 2007.
- MacMaster, Gordon, "Sign Language Translation Using Machine Learning and Computer Vision" (2020).UVM Honors College Senior Theses. 480.

To be Continued...