G H Patel College Of Engineering And Technology

Action Recognition System

Armin Vaghasiya(12002080501005)
Ayushi Patel(12002080501006)

Guided by
Prof. Rahul Patel

Objective

Our project aims to create a computer application and train a model which when shown a real time video of hand gestures of American Sign Language shows the output for that particular sign in text format on the screen. This technology can be used for a variety of applications, such as shopping mall.

Introduction

Sign Language Communicating language used primarily by deaf people.

Uses different medium such as hands, face, or eyes rather than vocal tract or ears for communication purpose. Communication using sign language.

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.

Literature Review

In recent years there has been tremendous research done on hand gesture recognition. With the help of literature survey done we realized the basic steps in hand gesture recognition are :-

1. Data acquisition
2. Data preprocessing
3. Feature extraction
4. Gesture classification

Software and Tools

1. Python
2. Tensorflow
3. OpenCV
4. NumPy
5. Matplotlib
6. Keras
7. PIL

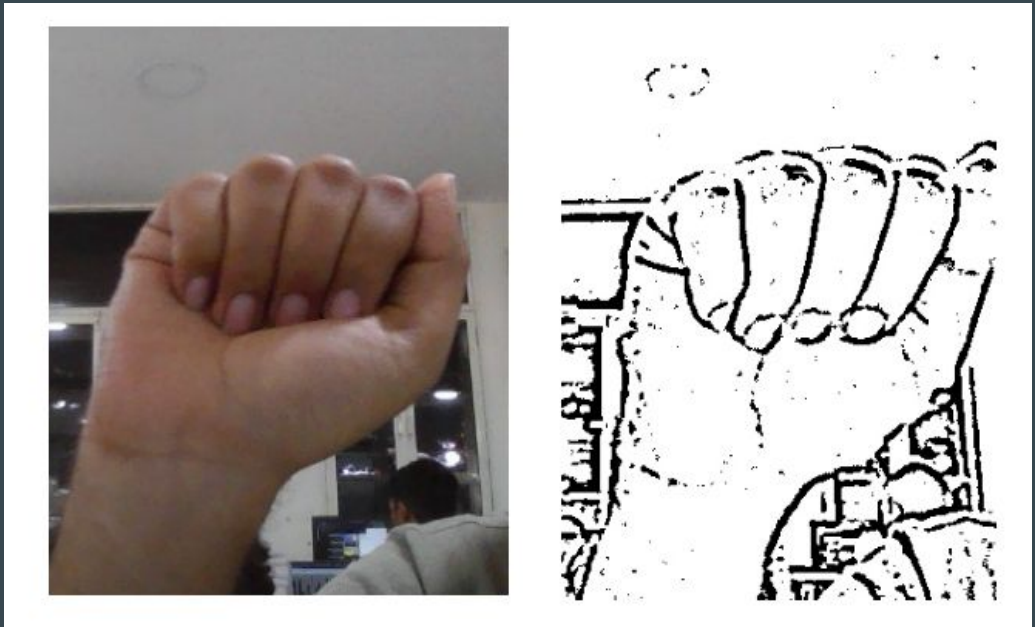
Methods

In recent years there has been tremendous research done on hand gesture recognition. With the help of literature survey done we realized the basic steps in hand gesture recognition are :-

1. Data acquisition
2. Data preprocessing
3. Feature extraction
4. Gesture classification

Implementation

1. Data Set Generation



Implementation

2. Image classification:

- CNN Model
- Activation Function (ReLu)
- Pooling Layer
- Dropout Layers
- Optimizer

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 126, 126, 32)	320
max_pooling2d (MaxPooling2D)	(None, 63, 63, 32)	0
conv2d_1 (Conv2D)	(None, 61, 61, 32)	9248
max_pooling2d_1 (MaxPooling2D)	(None, 30, 30, 32)	0
flatten (Flatten)	(None, 28800)	0
dense (Dense)	(None, 128)	3686528
dropout (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 96)	12384
dropout_1 (Dropout)	(None, 96)	0
dense_2 (Dense)	(None, 64)	6208

...
Total params: 3,716,378
Trainable params: 3,716,378
Non-trainable params: 0

Implementation

3. Video classification

Layer (type)	Output Shape	Param #
lstm_3 (LSTM)	(None, 60, 64)	442112
lstm_4 (LSTM)	(None, 60, 128)	98816
lstm_5 (LSTM)	(None, 64)	49408
dense_3 (Dense)	(None, 64)	4160
dense_4 (Dense)	(None, 32)	2080
dense_5 (Dense)	(None, 3)	99

Total params: 596,675
Trainable params: 596,675
Non-trainable params: 0

Conclusion

1. In this report, a functional real time vision based american sign language recognition for D&M people have been developed for asl alphabets.
2. We achieved final accuracy of 95.0% and 34% on our respective image and video datasets.

Limitations

1. The model works well only in good lighting conditions.
2. Plain background is needed for the model to detect with accuracy.
3. Video model is not accurate.
4. Proper GUI is required.

Future Scope

1. We are planning to achieve higher accuracy even in case of complex backgrounds by trying out various background subtraction algorithms.
2. We are also thinking of improving the preprocessing to predict gestures in low light conditions with a higher accuracy.

Thank You