Gesture Recognition On American Sign Language Employing Image Depth-Map Training On Videos

Amit Jaggi
2015013
amit15013@iiitd.ac.in

Mayank Mohindra
2015056
mayank15056@iiitd.ac.in

I. Data-set Used

A. MNIST Data-set

Mnisttraindata
Labeled training pixel data in CSV for 28x28 gray-scale (0-255) values of American Sign Language letters as hand shapes (27,455 labeled cases)
Mnsittestdata
Labeled testing pixel data in CSV for 28x28 gray-scale (0-255) values of American Sign Language letters as hand shapes (7172 labeled cases)

B. Finger-spelling Data-set

The data-set comprises 24 static signs (excluding letters j and z because they involve motion). This was captured in 5 different sessions, with similar lighting and background.It was used in the paper-Spelling It Out: RealTime ASL Finger-spelling Recognition.

II. Tasks Completed

- Data Acquisition
- Image Pre-processing
- Segmentation
- Feature Extraction
- Classification

A. Image Pre-processing and Segmentation

Transformation

1) Mnist dataset included grey-scale images which can be directly passed to the PCA.
2) All the depth images have cropped out to return only the hand part.
3) The coloured images have different sizes,so to bring similarity all the colored images have been re-sized to (73,128) pixels

B. Feature Extraction

For the extraction of the features, we use PCA to analyze on the important features and work optimally. Our main aim is to apply this training on videos.

C. Classification

The classifiers used for testing include: *(Non-Segmented — Segmented)*

1) Gaussian Naive Bayes (74.34% — 86.46%)
2) Random Forest (97.60% — 97.72%)
3) Decision Tree Classifier (90.19% — 92.84%)

III. Results Obtained

The accuracy is obtained for each on two data-sets:

1) On RGB images
2) On Segmented Images

As expected in all cases of classification, the accuracy for the segmented data-set is far more than without segmentation. In case of Gaussian NaiveBayes, it is as high as 11% more. This reveals the importance of Segmentation for ASL gestures.

IV. Tasks Remaining

A. Cropping Image — Real Noise Reduction

Creating a Hand/Non-Hand Binary Classifier which separates the required Hand image from the rest of the image.
The classifier randomly generates sliding window and returns the HOG vector of these sliding window along with the label of box: 1 for Hand and 0 for non-hand.

B. Alternative for Depth-Map Images

Histogram of Oriented Gradients(HOG) Technique is used for the purpose of object detection.The HOG vector is associated with each Sliding window in the hand/non-hand classifier which finally return the probability of the cropped image being a hand or not.

C. Smart Application to Videos

To realize our aim, we plan the application of Natural Language Processing (NLP) to consider the dependence of letters into likelihood estimation for efficient prediction of the meaning of the data.