Sign Language Recognition

Mukesh Makwana & Rathna G. N
Indian Institute of Science
Electrical Engineering

Contact Information:
DSP Lab
Electrical Engineering
Indian Institute of Science
Bangalore, Karnataka, India

Phone: +91 8880 994668
Email: mux032@gmail.com

Introduction

Communication between hearing-impaired minority and non-impaired majority is difficult as most of the time latter community is not aware of sign language.

Sign language consists of two major components namely Finger-spelling and Word level sign vocabulary. Finger-spelling is used to spell words letter by letter, whereas Word level sign vocabulary comprises of signed words and is used for the majority of communication.

Finger-spelling recognition can be broadly divided into two parts, extraction of feature vectors and classification of gestures using feature vectors. The two dataset used both contains depth images of gestures. The use of depth images eases the task of preprocessing and helps obtain results in real-time.

Datasets

First dataset used is ASL dataset collected by Byeongkeun *et al.* contains 31,000 images from five different subjects for 31 different gestures from each.

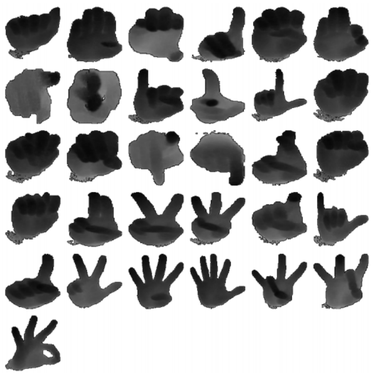

Figure 1: ASL Dataset (Depth Images)

Second dataset is ISL dataset which has been collected by us, it contains 630 images from six different subjects for five gestures from each.

Figure 2: ISL Dataset (Depth Images)

Methods

Few methods which are used for ASL dataset and ISL dataset are:

- **Random Forest:** Its an estimator that operates on a number of decision tree classifiers on various subsamples (using bootstrap aggregating) of the dataset.

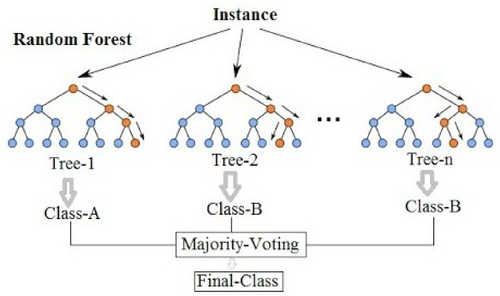

Figure 3: Random Forest

- **SVM:** SVM finds an optimal hyperplane between different classes. Principal Component Analysis (PCA) is applied before SVM model for dimensionality reduction.

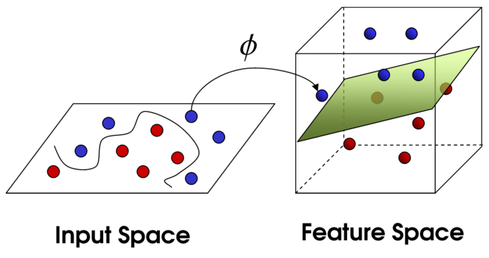

Figure 4: SVM for two classes

- **CNN:** Convolutional Neural Network is a deep neural network which uses the property of convolution. One important advantage of CNN over ordinary NN is the reduction in the number of parameters to be learnt. A CNN primarily consists of convolution layer, ReLU layer and Pooling layer. ReLU is an activation function layer. Pooling performs the down sampling operation and reduces the volume of output of ReLU.

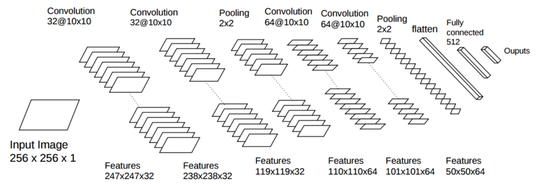

Figure 5: Convolutional Neural Network

Results

For ASL Dataset, we trained our models on four subjects and tested on fifth subject . And after taking all combination of subjects we get the average accuracy value. Accuracies are noted below in table,

Method	# of Sub.	# of class	Accur.(%)
SVM (4/1)	5	31	48.01
RF (4/1)	5	31	70
CNN (4/1)	5	31	70±1

* Diagonal of Confusion matrix plot shows the number of gestures which are perfectly classified (darkest in diagonal) and those which are confused with other gestures (faint in diagonal).

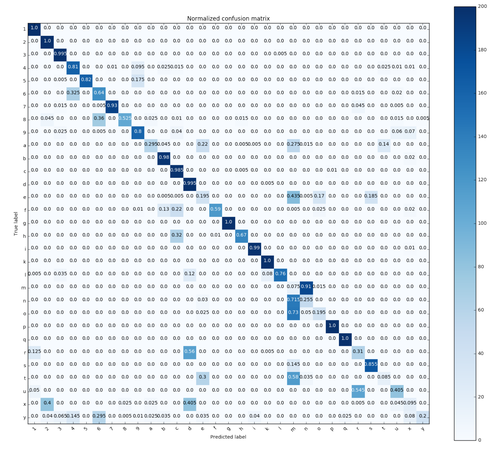

Figure 6: Confusion Matrix

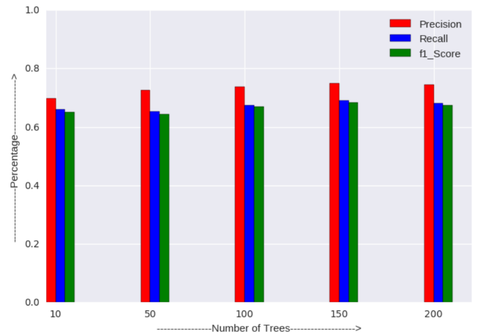

Figure 7: Random Forest with different numbers of trees

For ISL dataset, we trained our models on five subjects and tested on 6th subject. And after taking all combination of subjects we get the average accuracy value. Accuracies are noted below in table,

Method	# of Sub.	# of class	Accur.(%)
SVM (5/1)	6	5	98
RF (5/1)	6	5	68.7
CNN (5/1)	6	5	83.7

Conclusions

- SVM and Random forest are better with training time and accuracy, but there is a very little to no room for further improvement and it cannot be trained dynamically.
- Whereas CNN takes much time to train and requires a large dataset for efficiency. But can be trained on new dataset on-the-go. Further tweaking properly, might result in better accuracy than achieved so far.

Forthcoming Research

- Significant improvement must be made in resolving the ambiguity in detection of visually similar alphabets. for example a, e, m, n, o, s, t.
- The problem statement can be extended to word detection using fingerspelling in real-time.

References

1. Byeongkeun Kang, Subarna Tripathi, and Truong Q. Nguyen. Real-time Sign Language Fingerspelling Recognition using Convolutional Neural Networks from Depth map. Department of Electrical and Computer Engineering, San Diego, 2015.