SIGN LANGUAGE GESTURE RECOGNITION

Kaustav Vats, Lakshya Bansal, Deepanshu Badshah

Indraprastha Institute of Information Technology, Delhi

Problem Statement

This project aims to compare and analyze the results and accuracies of different classifiers(SVM, RF, LR, CNN etc.) with different feature extraction technique for hand gesture recognition.

Literature Review

The current ways, to perform hand gesture and sign language recognition mainly focus on pre-processing the data and applying CNN, ANN and SVM with different changes for recognition. The following are the papers which we are referring for our project.

- Recognizes using various feature extraction techniques like shape descriptors, SIFT and HOG individually alongwith SVM classifier.[1]
- Uses CNN (max pooling strategy) with dropout to classify images of both the the letters and digits in American Sign Language.[2]
- uses CNN with stochastic pooling strategy for classification of gestures in selfie videos.[3]

Dataset Description

- **LSA64:** Dataset for Argentinian Sign Language 3200 videos with 5 repetitions of 64 different type of signs. Database recorded in two sets.
  - 23 one-handed signs
  - 41 signs (22 two-handed and 19 one-handed).
- **Sign Language MNIST Kaggle Dataset** [0]
  - Train set - 27455 American Sign Language labeled grayscale images of size 28x28
  - Test set - 7172 labeled grayscale images of size 28x28

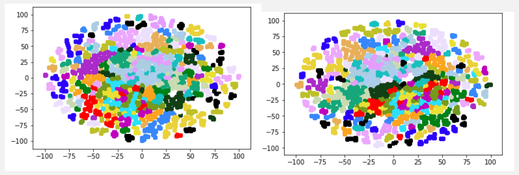
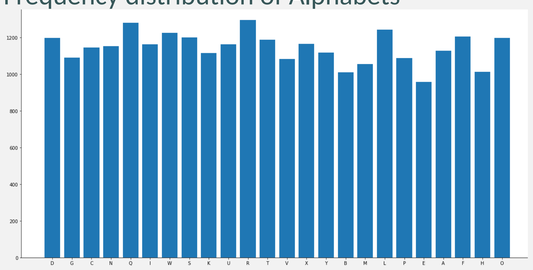


Figure: TSNE Plots for original data (left) and PCA reduced data(right)

Data Visualization

- Frequency distribution of Alphabets



- Data figure

Interpretation of Results

- It can be noted that apply PCA to the data results in an accuracy greater than or equal to the accuracy obtained when the original data is classified. Hence, the projected data has better variance of data as can be seen in 1
- Applying LDA (20 linear discriminators) increases the accuracy in 3 out of 4 (except Logistic Regression) classifiers used .
- Applying LDA after PCA decreases the accuracy of Random forest and Logistic Regression, meaning that the inter class distance in the resulting projection of the data is worse than with PCA.

Proposed Algorithm

Feature Extraction:

1. For Argentinian sign language dataset we have implemented bag of words model over the extracted frames of the training videos and the histograms are computed.
2. For MNIST data set We have used sift and HOG techniques for the feature extraction and obtained images data with lesser number of features.

Sign Recognition:

1. For Argentinian sign language dataset we have trained SVM over the training histograms computed above and classification is done on the test histograms computed from the test videos.
2. For MNIST dataset we have used different classifiers(SVM, RF, LR and CNN) for classification.

Results

The following results were obtained on the mnist dataset of hand gestures using various classifiers with HoG feature extractor.

	N components	Training Accuracy (%)	Testing Accuracy (%)
PCA	192 (no reduction)	81.23	61.85
	180	91.90	73.25
	100	91.45	76.28
LDA	1	21.12	19.38
	10	84.72	74.39
	20	93.81	82.75
	23 (upper bound)	95.01	81.88
PCA → LDA	100 (PCA), 20 (LDA)	90.12	78.20
LDA → PCA	20 (LDA), 10 (PCA)	84.72	73.29

	N components	Training Accuracy	Testing Accuracy
PCA	192 (no reduction)	98.77	87.15
	180	98.77	87.06
	100	98.35	86.01
LDA	1	20.30	15.85
	10	86.90	74.37
	20	97.15	87.16
	23 (upper bound)	98.10	81.92
PCA → LDA	100 (PCA), 20 (LDA)	94.50	77.76
LDA → PCA	20 (LDA), 10 (PCA)	90.12	78.20

	N components	Training Accuracy	Testing Accuracy
PCA	192 (no reduction)	100	91.23
	180	100	89.36
	100	100	90.40
LDA	1	99.91	15.01
	10	100	76.85
	20	100	84.31
	23 (upper bound)	100	83.11
PCA → LDA	100 (PCA), 20 (LDA)	100	85.52
LDA → PCA	20 (LDA), 10 (PCA)	100	77.14

	N components	Training Accuracy	Testing Accuracy
PCA	192 (no reduction)	84.70	73.43
	180	85.43	74.14
	100	89.73	80.34
LDA	1	24.10	19.32
	10	98.61	79.71
	20	99.93	87.11
	23 (upper bound)	99.96	88.44
PCA → LDA	100 (PCA), 20 (LDA)	99.10	83.40
LDA → PCA	20 (LDA), 10 (PCA)	98.67	79.82

Figure: Accuracies obtained by using Gaussian Naive Bayes (top-left), Logistic Regression (top-right), Random Forest(bottom-left) and SVM (bottom-right).

Conclusion

- HOG features were performing better than the Bag of visual words.
- Random forest's accuracy was highest with complete features(Accuracy was reduced after PCA and LDA)

References

[1] Juhi Ekbote, Mahasweta Joshi, âĂIJIndian Sign Language Recognition using SVM and ANN classifiersâĂİ

[2] Vivek Bheda, N. Dianna Radpour, âĂIJUsing Deep Convolutional Networks for Gesture Recognition in American Sign LanguageâĂİ

[3] G.Anantha Rao, K.Syamala, P.V.V.Kishore, âĂIJDeep Convolutional Neural Networks for Sign Language RecognitionâĂİ

[4] Kaggle Mnist Dataset https://www.kaggle.com/datamunge/sign-