ME 592X

Assignment 3

Engineering Image Analysis – Group 1 – Combustion Image Set

Given Data Set:

- Training Set – 54000 x 25000
- Testing Set – 18000 x 25000
- Validation Set – 9000 x 25000

Code: Run combustion-classifier.ipynb (Python notebook) or combustion-classifier.py (Python code). Please change path to the combustion data if extracting data from different location (Second cell in Python notebook).

Preprocessing:

- The flattened images are reshaped into (100x250).

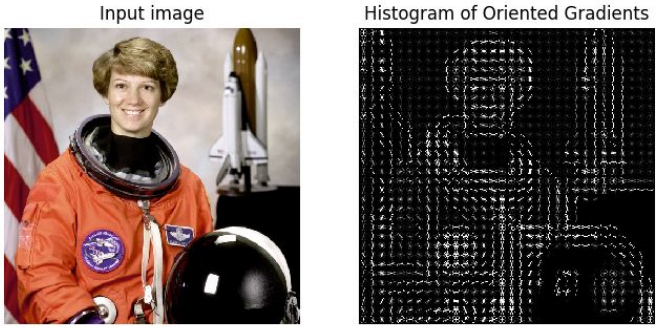
Features Extracted:

- The given Combustion image set does not contain any clear edges, because the images are of flames and smoke - don't contain any solid object.
- Methods like Canny edge detector doesn't capture edges correctly.
- Thus, the following two features are used for the feature extraction: HoG, Dictionary learning

i) HOG – Histogram of Oriented Gradients

- HoG counts occurrences of gradient orientation in localized portions of an image
- https://www.learnopencv.com/histogram-of-oriented-gradients/
- http://scikit-image.org/docs/dev/auto_examples/features_detection/plot_hog.html

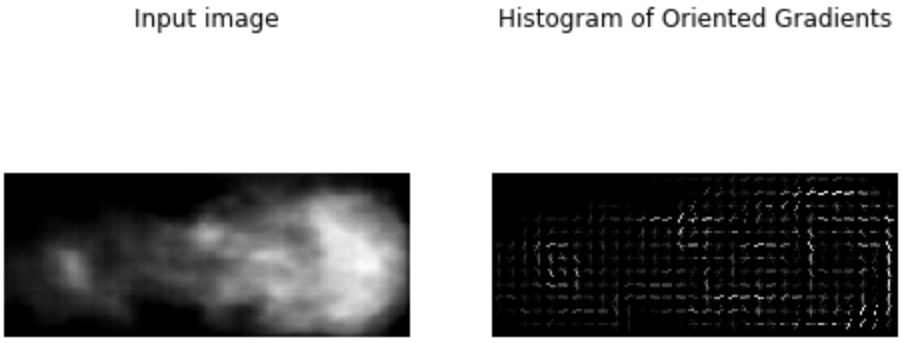
Ex:



- Changing the following parameters of the Hog alters the reduction in dimension – The trade-off between the capturing the feature and dimension reduction
    - Orientation
    - Pixels per cell (dimensions of each block)

Orientation	Pixels Per Cell	Feature Vector Dimension
5	10x10	1250
3	16x16	270

Sample Image (HoG)



ii) Dictionary learning

- Constructs a dictionary that can be used to represent data using a sparse code.
- Requires as few as 10 features for the given dataset, for >90% accuracy of classification, hence significantly reducing dimensionality of the dataset.
- Reference code: http://scikit-learn.org/stable/modules/generated/sklearn.decomposition.MiniBatchDictionaryLearning.html#sklearn.decomposition.MiniBatchDictionaryLearning

Random Forest

- **http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html**
- Following parameters are altered to find out the nature of impact they have on the accuracy
    - Number of trees (n_estimators)

- Number of features considered at each node (n_features)
- Number of layers (max depth)

Number of Trees	Max Depth	Number of Features	Accuracy
2	**2**	**32**	**0.875055555556**
2	**4**	**64**	**0.945055555556**
6	**2**	**128**	**0.884444444444**
6	**4**	**1250**	**0.971388888889**

Dictionary Learning

Number of Trees	Max Depth	Number of Features	Accuracy
2	**2**	**4**	**0.830277777778**
2	**4**	**4**	**0.919722222222**
6	**2**	**8**	**0.819777777778**
6	**4**	**10**	**0.913055555556**

Boosting Techniques

Boosting Technique	Accuracy with HoG features	Accuracy with Dictionary learning
Decision Trees	**0.980944444444**	**0.980388888889**
Extra Tree	**0.998277777778**	**0.993611111111**
Gradient Boosting	**0.997777777778**	**0.910722222222**
Adaboost	**0.998055555556**	**0.960055555556**
Voting Classifier	**0.997166666667**	**0.990833333333**

Tuning Hyper Parameters

K-fold cross validation with Grid Search

- Range of the hyper Parameters considered for grid search

For Random Forest classifier:

Hyper Parameter	**Set**
No of Trees	**{2,3}**
Max Depth	**{2,4}**
Max Features	**{10.30}**

Final Hyper Parameter

(screenshot of the grid search output)

Tuning hyper-parameters for precision

Best parameters set found on development set, **using HOG features**:

{'max_features': 30, 'n_estimators': 3, 'max_depth': 2}

Grid scores on development set, using HOG features:

0.877 (+/-0.167) for {'max_features': 10, 'n_estimators': 2, 'max_depth': 2}
0.924 (+/-0.179) for {'max_features': 10, 'n_estimators': 3, 'max_depth': 2}
0.915 (+/-0.183) for {'max_features': 30, 'n_estimators': 2, 'max_depth': 2}
0.924 (+/-0.192) for {'max_features': 30, 'n_estimators': 3, 'max_depth': 2}
0.907 (+/-0.172) for {'max_features': 10, 'n_estimators': 2, 'max_depth': 4}
0.913 (+/-0.175) for {'max_features': 10, 'n_estimators': 3, 'max_depth': 4}
0.917 (+/-0.187) for {'max_features': 30, 'n_estimators': 2, 'max_depth': 4}
0.922 (+/-0.192) for {'max_features': 30, 'n_estimators': 3, 'max_depth': 4}

Detailed classification report:

The model is trained on the full development set, using HOG features.
The scores are computed on the full evaluation set, using HOG features.

	precision	recall	f1-score	support
0	0.99	0.81	0.89	15000
1	0.81	0.99	0.89	12000
avg / total	0.91	0.89	0.89	27000

 # Tuning hyper-parameters for recall

Best parameters set found on development set, **using HOG features**:

{'max_features': 10, 'n_estimators': 3, 'max_depth': 4}

Grid scores on development set, using HOG features:

0.882 (+/-0.163) for {'max_features': 10, 'n_estimators': 2, 'max_depth': 2}
0.843 (+/-0.227) for {'max_features': 10, 'n_estimators': 3, 'max_depth': 2}
0.864 (+/-0.192) for {'max_features': 30, 'n_estimators': 2, 'max_depth': 2}
0.882 (+/-0.257) for {'max_features': 30, 'n_estimators': 3, 'max_depth': 2}
0.879 (+/-0.240) for {'max_features': 10, 'n_estimators': 2, 'max_depth': 4}
0.899 (+/-0.252) for {'max_features': 10, 'n_estimators': 3, 'max_depth': 4}
0.885 (+/-0.256) for {'max_features': 30, 'n_estimators': 2, 'max_depth': 4}
0.896 (+/-0.273) for {'max_features': 30, 'n_estimators': 3, 'max_depth': 4}

Detailed classification report:

The model is trained on the full development set, using HOG features.
The scores are computed on the full evaluation set, using HOG features.

	precision	recall	f1-score	support
0	0.96	0.63	0.76	15000
1	0.67	0.97	0.79	12000
avg / total	0.83	0.78	0.77	27000

 # Tuning hyper-parameters for precision

Best parameters set found on development set, using **Dictionary learned features:**

{'max_features': 10, 'n_estimators': 2, 'max_depth': 4}

Grid scores on development set, using Dictionary learned features:

0.805 (+/-0.059) for {'max_features': 3, 'n_estimators': 2, 'max_depth': 2}
0.748 (+/-0.148) for {'max_features': 3, 'n_estimators': 3, 'max_depth': 2}

0.793 (+/-0.174) for {'max_features': 10, 'n_estimators': 2, 'max_depth': 2}
0.840 (+/-0.167) for {'max_features': 10, 'n_estimators': 3, 'max_depth': 2}
0.867 (+/-0.128) for {'max_features': 3, 'n_estimators': 2, 'max_depth': 4}
0.834 (+/-0.291) for {'max_features': 3, 'n_estimators': 3, 'max_depth': 4}
0.889 (+/-0.166) for {'max_features': 10, 'n_estimators': 2, 'max_depth': 4}
0.880 (+/-0.159) for {'max_features': 10, 'n_estimators': 3, 'max_depth': 4}

Detailed classification report:

The model is trained on the full development set, using Dictionary learned features.
The scores are computed on the full evaluation set, using Dictionary learned features.

	precision	recall	f1-score	support
0	0.91	0.92	0.92	15000
1	0.90	0.89	0.89	12000
avg / total	0.91	0.91	0.91	27000

 # Tuning hyper-parameters for recall

Best parameters set found on development set, using **Dictionary learned features:**

{'max_features': 10, 'n_estimators': 2, 'max_depth': 4}

Grid scores on development set, using Dictionary learned features:

0.754 (+/-0.120) for {'max_features': 3, 'n_estimators': 2, 'max_depth': 2}
0.747 (+/-0.161) for {'max_features': 3, 'n_estimators': 3, 'max_depth': 2}
0.799 (+/-0.235) for {'max_features': 10, 'n_estimators': 2, 'max_depth': 2}
0.729 (+/-0.121) for {'max_features': 10, 'n_estimators': 3, 'max_depth': 2}
0.830 (+/-0.198) for {'max_features': 3, 'n_estimators': 2, 'max_depth': 4}
0.862 (+/-0.199) for {'max_features': 3, 'n_estimators': 3, 'max_depth': 4}
0.869 (+/-0.243) for {'max_features': 10, 'n_estimators': 2, 'max_depth': 4}
0.846 (+/-0.222) for {'max_features': 10, 'n_estimators': 3, 'max_depth': 4}

Detailed classification report:

The model is trained on the full development set, using Dictionary learned features.
The scores are computed on the full evaluation set, using Dictionary learned features.

	precision	recall	f1-score	support
0	0.91	0.92	0.92	15000

1	0.90	0.89	0.89	12000
avg / total	0.91	0.91	0.91	27000

Comparison of Ensemble Classifier

HoG:

Ensemble Classifier	Optimal Parameters		
	Max_features	N_estimators	Max_depth
Random Forest	**30**	**3**	**4**
Decision Tree	**30**		**4**
Extra Trees	**30**		**4**
Ada Boost		**3**	
Gradient Tree Boosting	**30**	**3**	**4**

Dictionary Learning:

Ensemble Classifier	Optimal Parameters		
	Max_features	N_estimators	Max_depth
Random Forest	**9**	**2**	**4**
Decision Tree	**9**		**4**
Extra Trees	**9**		**4**
Ada Boost		**3**	
Gradient Tree Boosting	**9**	**3**	**4**

Accuracy of boosting techniques after grid search (Optimal parameters):

Boosting Technique	Accuracy with HoG features	Accuracy with Dictionary learning
Decision Trees	**0.924833333333**	**0.9125**
Extra Tree	**0.889**	**0.939222222222**
Gradient Boosting	**0.996111111111**	**0.978**
Adaboost Classifier	**0.885888888889**	**0.874777777778**