# FINAL PROJECT: CLASSIFICATION OF VISUAL INFORMATION

*Karishma Ganesh Reddy and Harshdeep Singh Chawla*

Rochester Institute of Technology

## 1. INTRODUCTION

The objective of this project is to utilize the knowledge obtained from the course throughout the semester to develop a classifier. This project proposes a classifier that is capable of recognizing the class or category of an object (in terms of whether it is a face, motorcycle or an airplane) present in an image. The classifier is trained on three different classes of objects namely airplanes, faces and motorcycles, using SIFT and HOG feature descriptors extracted from the images of all the classes. In the final step, the accuracy of the classifier is obtained by evaluating the performance of the classifier when images other than those used for training are provided to the classifier. The evaluation is in terms of whether the object in the image is properly classified or not.

## 2. METHODS

The image dataset for conducting the experiment is partitioned to form the testing set and the training set. No image is common to both the sets so as to ensure that an accurate classifier is developed which is capable of handling data from outside the training set as well.

The next step is to perform pre-processing of the images present within these sets, so that the variation in terms of illumination, scale etc. present in these images is eliminated. The scale variation is eliminated by fixing the same size (aspect ratio) for all the images in the image set. This process is carried out only before extracting HOG features and not before extracting SIFT features because SIFT features are scale invariant. Preprocessing for illumination invariance is carried out by performing normalization on all the images before extracting their HOG or SIFT feature descriptors. All the images are also converted to their gray scale single precision equivalents as required by the library functions that will be extracting the features from these images.

Next, the SIFT and the HOG features are extracted from the images of the sets using their respective library functions. For the case of SIFT features, these features are further processed to create a 'Visual Words vocabulary', more popularly referred to as the 'Bag of Words'. This 'Visual Words' vocabulary is created by initially grouping similar features together to form clusters. Each cluster so obtained is compact and distinct from other clusters and represents a visual word. Next, the feature vector for the images are created using the visual words obtained from the previous process. This is done by counting the occurrences of each visual word within that image and developing a histogram accordingly which becomes the feature vector for that image (new reduced representation of the image).

For the case of HOG features, the features are extracted using a cell size of 16 pixels and the blocks formed consisted of 4 such cells. The HOG feature vector is arranged by these blocks. The fixed size for all the images ensures a feature vector of same length for all the images.

Feature conditioning was not required for the SIFT feature vectors since the processing of these features to form the 'Bag of Words' as explained before ensures that the features are robust and discriminative and also reduces the dimensions. In the case of extracted HOG features, since the library function being used for classification required a fixed length feature size, no dimensionality reduction processing was carried out.

The feature vectors obtained from the images form the foundation on which the training of the classifier is carried out. The next step involves creation of two matrices, one holding the feature vectors of the training set and the other holding the feature vectors of the testing set for each kind of feature vector (SIFT and HOG). Labels are also provided to the training set feature vector matrix as well as the testing set feature vector matrix for each type of feature vector. The training and testing of the classifier is carried out separately for each kind of feature vector using the corresponding feature vector matrices in which the training set feature vector matrix along with the training set labels is used for training the classifier and the testing set feature vector matrix along with the testing set labels is used for testing the classifier.

The classifier being used for this project is the SVM classifier. The training of this classifier is carried out by first dividing the image set into 4 equal sections. Each section contains equal proportion of images from each of the three categories. The classifier is first trained on the training set made up of 3 sections of the image set and tested on the testing set consisting of a single section of the image set that was not used for training. Next, the classifier is trained on 2 sections of the image set present in the training set and tested on the other 2 sections present within the testing set that were not used for training. Finally, the classifier is trained on 1 section of the image set in the training set and tested on the other 3 sections present within the testing set that were not used for training.

The entire process mentioned above is carried out separately for each kind of feature vector and the accuracy for each scenario is provided in Table 1 which is displayed in the results section.

## 3. EXPERIMENTAL SETUP

The images used for the project were taken from the Caltech 101 Object Categories Dataset. The image set was formed by taking equal number of images from each category i.e. Face, Motorbike and Airplane datasets of the Caltech 101 Object Categories Dataset.

Preprocessing for the images is carried out by resizing the images using imresize function of MATLAB before extracting HOG features to eliminate scale variance between images. To eliminate illumination variance, normalization of the images was done before extraction of both SIFT features as well as the HOG features and was carried out using the mat2gray MATLAB library function.

For feature extraction, vl_sift function from the vlfeat library was used for extracting SIFT features and Matlab function extractHOGfeatures was used for extraction of HOG features vectors.

For the extracted SIFT features, the visual vocabulary or the 'bag of words' is obtained by making use of Matlab's bagOfFeatures function on the obtained features. The bagOfFeatures function provides a method called encode using which we can create the feature vector for each image. This feature vector consists of a histogram describing the number of occurrences of each visual word in the corresponding image.

Next, the HOG and the SIFT feature vectors are provided to their respective SVM classifiers. For the SIFT feature vectors, the SVM classifier is implemented by using the functions from the LIBSVM library. The function 'svmtrain' is used for training the classifier on the training dataset using the training labels and the feature vectors of the images present within the training set provided to this function and the function 'svmpredict' is used for testing the classifier on the testing dataset using the testing labels and the feature vectors of the images present within the testing set and the trained classifier model provided to this function. The svmpredict function also provides the accuracy of the classifier's results along with the predicted labels for the images in the test set.

For the case of the HOG feature vectors, the SVM classifier is implemented using the MATLAB function fitcecoc. The training set feature vectors and their labels are provided as inputs to this function. The labels for the images present within the test set is predicted using the predict function of MATLAB, which takes the test set feature vectors as input along with the trained classifier. These predicted labels along with the labels of the test set images are then provided to the confusionmat function of MATLAB that generates a confusion matrix, based on which the accuracy of the classifier is calculated.

## 4. RESULTS

The results obtained for each stage of training are displayed in the table as follows:

| | Number of sections of image set in the training set. | Number of sections of image set in the testing set. | Accuracy obtained. (SIFT) | Accuracy obtained. (HOG) |
|---|---|---|---|---|
| 1 | 3 | 1 | 96.6% | 97% |
| 2 | 2 | 2 | 95.4% | 96.2% |
| 3 | 1 | 3 | 95.1% | 95.6% |

Table 1: Displaying accuracy obtained for the cases considered for both HOG as well as SIFT features.

## 5. DISCUSSION

As is evident from the analysis of the results table above, we observe that as we decrease the amount of data present in the training set, the accuracy obtained for the corresponding classifier also decreases. This can be attributed to the fact that since the classifier is trained on less data, it could not learn much about the features associated with a particular category.

Also, we observe that the accuracy obtained for classification using HOG feature vectors is a little more compared to the accuracy obtained for classification using SIFT feature vectors since HOG feature vectors provide a global description of a given category in the sense that they are able to obtain uniqueness of each category from a global point of view.

## 6. REFERENCES

[1]http://www.vision.caltech.edu/html-files/archive.html

[2]http://www.vlfeat.org/

[3]http://www.mathworks.com/help/vision/ref/extracthogfeatures.html

[4]http://www.mathworks.com/help/vision/examples/image-category-classification-using-bag-of-features.html

[5]http://www.mathworks.com/help/vision/ref/bagoffeatures-class.html

[6]http://www.mathworks.com/help/vision/ref/trainimagecategoryclassifier.html

[7]https://www.csie.ntu.edu.tw/~cjlin/libsvm/

[8]http://www.mathworks.com/help/vision/examples/digit-classification-using-hog-features.html

[9]https://en.wikipedia.org/wiki/Histogram_of_oriented_gradients

[10]http://www.mathworks.com/help/stats/fitcecoc.html

[11]http://stackoverflow.com/

[12] http://www.mathworks.com/help/stats/confusionmat.html

[13]http://www.mathworks.com/help/stats/compactclassificationdiscriminant.predict.html