# Car Classification using Principal Components Analysis

# Principal Components Analysis

The objects of interest for this project were rear ends of cars. The data set consisted of 95 training images, and 31 testing images (all of the same category). The negative testing set consisted of 11 images form the Sample Pictures folder (Windows).

The number of eigenvectors were chosen heuristically, through empirical trial and error. The final number of subdimensions reached was 70.

Classification involved projecting both the positive and negative data sets onto the eigenspace. Both the data sets were then reconstructed back to the original space. There are at least two simple options for classification. The first is to measure the Euclidean distance from every image to the mean image in both the data sets, observing the max, min, and average difference values for each set. The second involves measuring the Euclidean distance between every image in the testing sets with every image/class in the training set. This is more computationally intensive, but will give more fine-grained statistics for the differences. A threshold value, theta, is then determined, above which we classify as non-car-rear-end and below which we classify as car-rear-end. The initial value for theta was the average value of the smallest difference for each data set.

Then, theta was varied between the extremes; the lowest difference in the positive data set, and the highest difference in the negative data set, with steps of 0.1e+03. The number of correct detections, misdetections, false alarms, and correct rejections were then recorded. A graph of false hits vs. hits was then plotted to give an ROC curve.
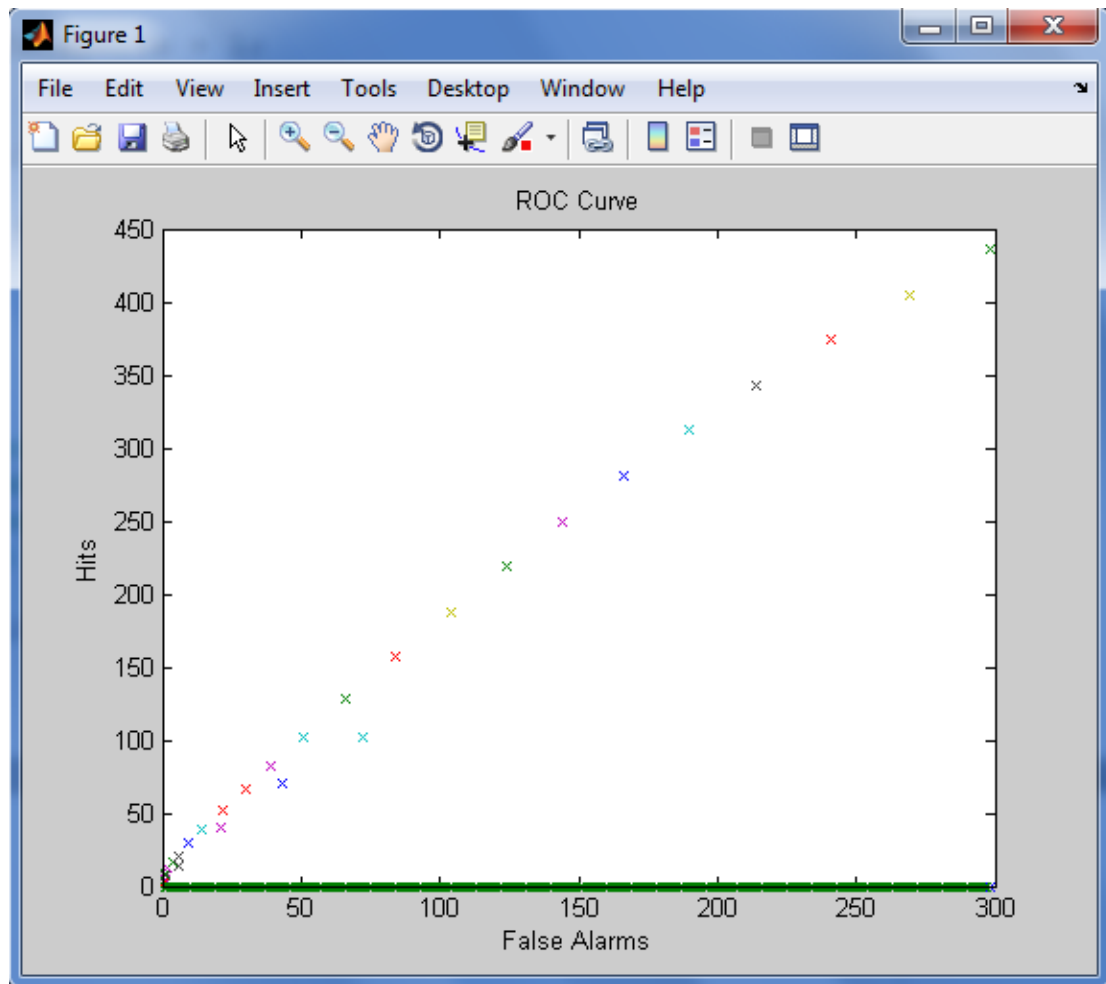
Note: Please see the attached source code for the bulk of the work