Ship Detection

Team Members
Henry Gridley
Haley Weinstein
Jon White

Programming Language
Matlab

Thesis
We want to investigate criminal activity in the City of Chicago to predict when and where law enforcement should focus their efforts when patrolling the city. We aim to predict areas where crime is most prevalent on a given day.

Data Source
https://www.kaggle.com/c/airbus-ship-detection/data

Description
1. Satellite images some of which contain ships, however there are some without and some that don't even contain water.
2. A CSV file containing locations of ships in the images if there are any encoded using run length encoding

Outline of Approach
We will be using color histograms, histograms of oriented gradients, and spatial information from each image as the features. These features will then be used to train an SVM classifier to detect ships. To perform detection, a sliding window will pass over each image and the features from that window will be extracted and used along with the trained model to predict whether or not there is a ship in that window. If so, a count will be added to a heatmap which can then be thresholded to remove erroneous classifications. Finally, bounding boxes will be drawn from the heatmap and overlaid on the image.

Our general approach will be:
1. Feature extraction
    a. Separate color channels of image and create a histogram from the values
        i. The 3 histograms will be appended together to create color feature vector
    b. Extract histogram of oriented gradients (HOG) features from image

    c. Resize image and convert to a column vector of spatial features
    d. Append all three feature vectors (color, HOG, spatial) to create full feature vector
2. Training
    a. Perform feature extraction on approximately 2000 ship examples and 2000 non-ship examples
    b. Use features to train SVM classifier
3. Detection
    a. Slide window across image
    b. Perform feature extraction on window and use to predict with trained SVM classifier model
    c. Keep pixel heatmap to know where “hits” are
    d. Threshold heatmap to remove incorrect classifications
        i. Should have a lower hit count than correct classifications
    e. Draw bounding boxes around ships based on heatmap