**CS 231A Final Project Report**
**Whale Face Identification: Can you distinguish these faces?**

**SeungWoo Jung**

**Abstract**
I have formulated North Atlantic Right Whale detector and classifier that is specifically used to localize whale from aerial photos, extract HoG feature, and train a set of binary SVM for identifying individual whale from the image. Localization algorithm uses spatial histogram distance comparison using Earth Mover's distance and median filtering to successfully capture a bounding box.

**Introduction**
North Atlantic Right Whales are one of the most endangered whale species in the world, currently being protected by U.S. Endangered Species Act. Most of the remaining 400 or so individuals live in North Atlantic ocean, where scientists are using helicopter aerial photos to keep track of each individuals' migration status. National Oceanic and Atmospheric Administration (NOAA)'s scientists are responsible for manually labeling each photo as specific individual, yet the work is often error prone and labor intensive. Objective of the project is to develop a framework where each whale's anatomical features can be extracted and used as an automated classification problem. This project is directly inspired by Kaggle Competition.

**Consideration of Images**
Anatomical feature of the right whale from aerial photo is limited, but there are some distinctive parasitic growths on dorsal side of each whale that differentiates one from another. Since there are no noticeable difference between color of whale's body and water, callosum shape is the only differentiating factor in whale "face." Different illumination also contributes to different water and whale color, barring successful localization. Another limiting factor is the difficulty in rectification because of no information about the distance between camera and object. Whale size cannot be assumed to be static because photos have been taken in the last 10 years.


**Figure 1**: Individual whales have different shapes and sizes of callosum on the center and/or sides of the head, and they might be continuous or broken.

**Previous Work**
Previous work by deepsense.io utilizes CNN for localization, alignment, and classification of the whales, by training head feature to bound a whale head. Deepsense.io has used differing depth of CNN for each task, while manually training each CNN with hand-labeled local features of whale such as blowhead and bonnet-tip. While CNN captures different layers of features from the object and therefore is superior in terms of preventing overfitting and

showing better performance, I wanted to apply general feature detectors and use traditional classifier to 1) automate the training process and 2) save the computation time.

**Technical Methodology**

Overall setup is a framework where each layer: 1) localize on whale head, 2) align the head to specific angle, and 3) extract feature from the cropped and rotated image to form feature that can be supplied to SVM for classification. There were 457 individuals that comprised each of the SVM classes, and there were 4,700 images for training set and 11,400 images for test set. Training set did not contain even number of images so it was difficult to generalize the rectification options.

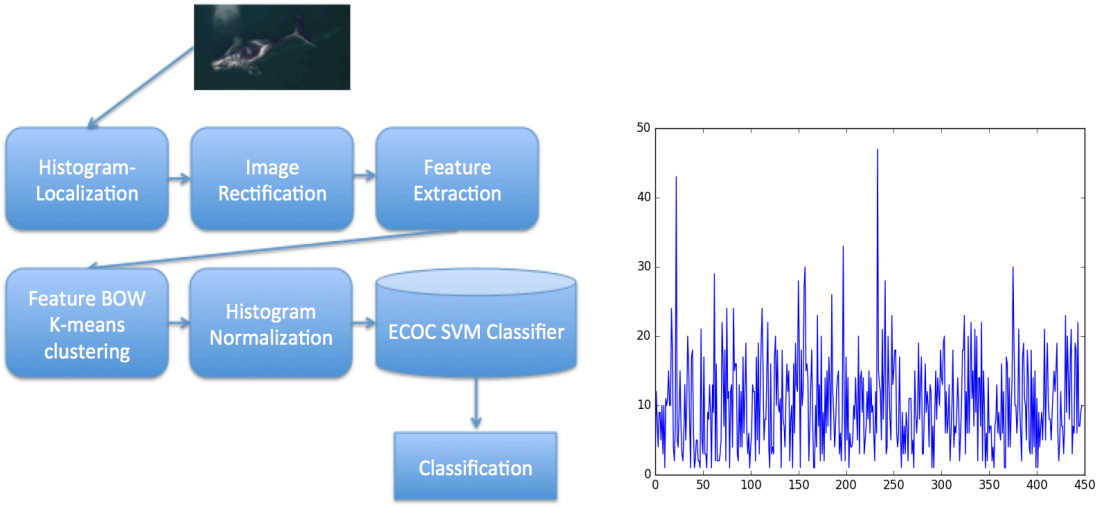


**Figure 2:** (left) overall workflow of image classification, (right) training set contained uneven occurrences of whale photos.

**1) Test of Feature Extraction**

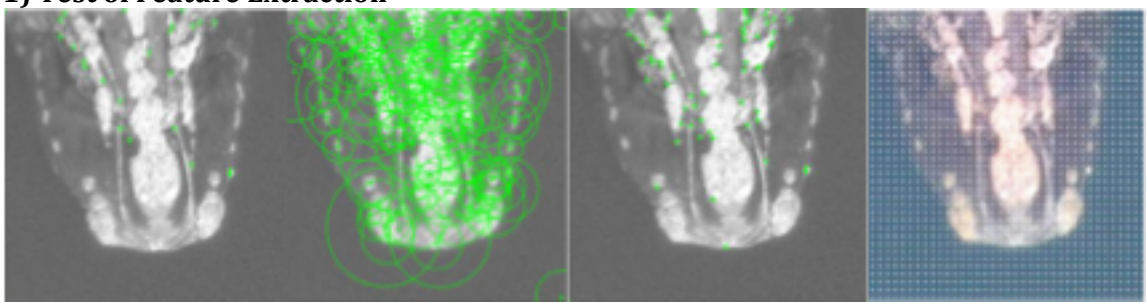

**Figure 3**: (from left to right) FAST, SURF, Harris, and HoG feature detection.

Before moving to localization of whale itself, I have experimented with descriptor algorithms that are specialized in different features of the image: FAST and Harris detectors for corners of the image, SURF a robust, blob detector, and HoG for gradient detection, which requires alignment. FAST and Harris detectors had feature space too sparse for correct histogram generation required for classification, so I had SURF and HoG as two candidates for feature detection.

**2) Localization**

Early attempts in localization involved clustering of descriptors based on their location and intensity, but soon I have realized that simple k-means or meanshift clustering did not quite catch the descriptors well. The reason was that there were too much wave and features clustered in non-head area as well.

The algorithm for localizing whale is as follows:

1) normalize each color channel, and compute median for entire matrix for each color channel.
2) conduct histogram equalization for added contrast
3) also construct histogram for each color channel matrix
4) divide image into subgrids (50x50 pixels) and compute histogram for each color channel
5) compute Earth Mover's Distance between entire image histogram and subgrid histogram for each color channel, and add distances.
   - Distance matrix D[x,y] = sum of emd_distances
   - size(D) = [ size(im, 1)/subgrid_size, size(im, 2)/subgrid_size ]
6) interpolate a function based on distance matrix and output an EMD distance image array.
7) add both and normalize to acquire a image heatmap that can be used to extract bounding box.

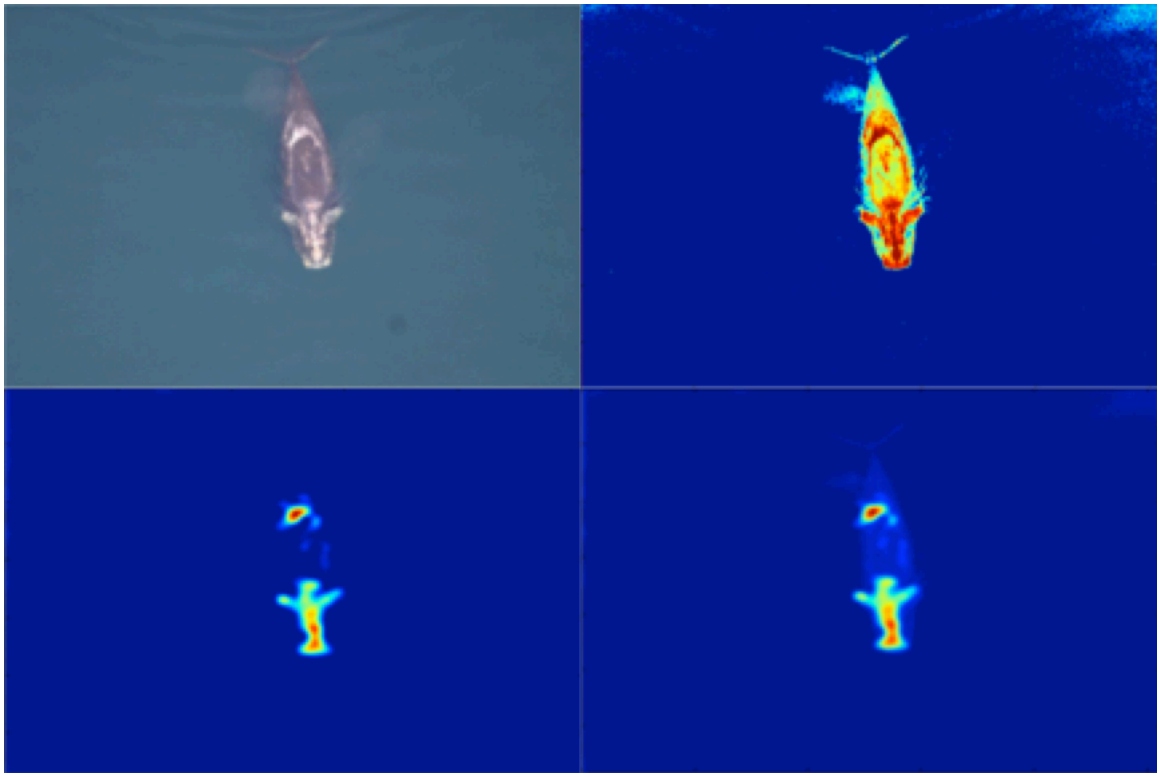

**Figure 4:** Various parts of localization algorithm. (Top-left) original image. (Top-right) after median subtraction and histogram equalization to enhance contrast. (Bottom-left) subgrid vs. whole image histogram distance map, based on Earth-Mover's Distance. (Bottom-right) Both images added and normalized.

I have realized that in detecting whale outline, I had to take into account the color difference and color distribution. Naturally, color difference can be accounted for using the median filter and histogram equalization since by using histogram equalization, image region with similar pixel intensities can be further distributed. Also, color distribution in a small enough area carries a context that should be utilized.

Basic assumption of the problem was that most of the image is water and not whale, and whale's dorsal callosity patterns can be summarized as a visual bag of words that can be transformed to histogram and can be compared. Last assumption was that those subgrid color patterns of whale and water will be distinguishable by Earth Mover's distance.

I have experimented with L2 norm and Earth Moving Distance (EMD) for a histogram distance calculation. I had a general intuition that Earth Moving distance is a metric based on how probable that both histograms have same distribution given both observations, but it has proven to be more accurate.

As a result of localization I got bounding box of the whale, from which I could calculate angle of rotation from the bounding box's diagonal line. I have rotated cropped image to align them the same.

**3) Classification Models**

I have used standard multiclass SVM (ECOC SVM) with default linear classifier. For training set I have extracted and aggregated HoG feature that I clustered using K-means clustering using num_center = 20. I have utilized problem set 4 spatial pyramid model to train and predict the ECOC SVM.

I calculated multiclass prediction accuracy using Log-loss = $-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}\log(p_{ij})$ for both SURF and HoG feature-based SVMs, where I have used a set of binary trainers that assigned $p_{ij}$ = 0 if ith test sample did not belong to class j, and $p_{ij}$ = 1 if it belonged to a class.

**Experimental Results**

Prediction accuracy for log-loss of SVM has been reported as 19.2 and 22.1 for HoG feature and SURF feature-based SVM. Since baseline for submission in Kaggle competition is 34.4 and winning team acquired 0.596, there has been little accuracy gained by my classification method. Possible reasons for such unsuccessful classification can be due to localization error, particularly in still distinguishing waves apart from whale callosity, and no rectification.

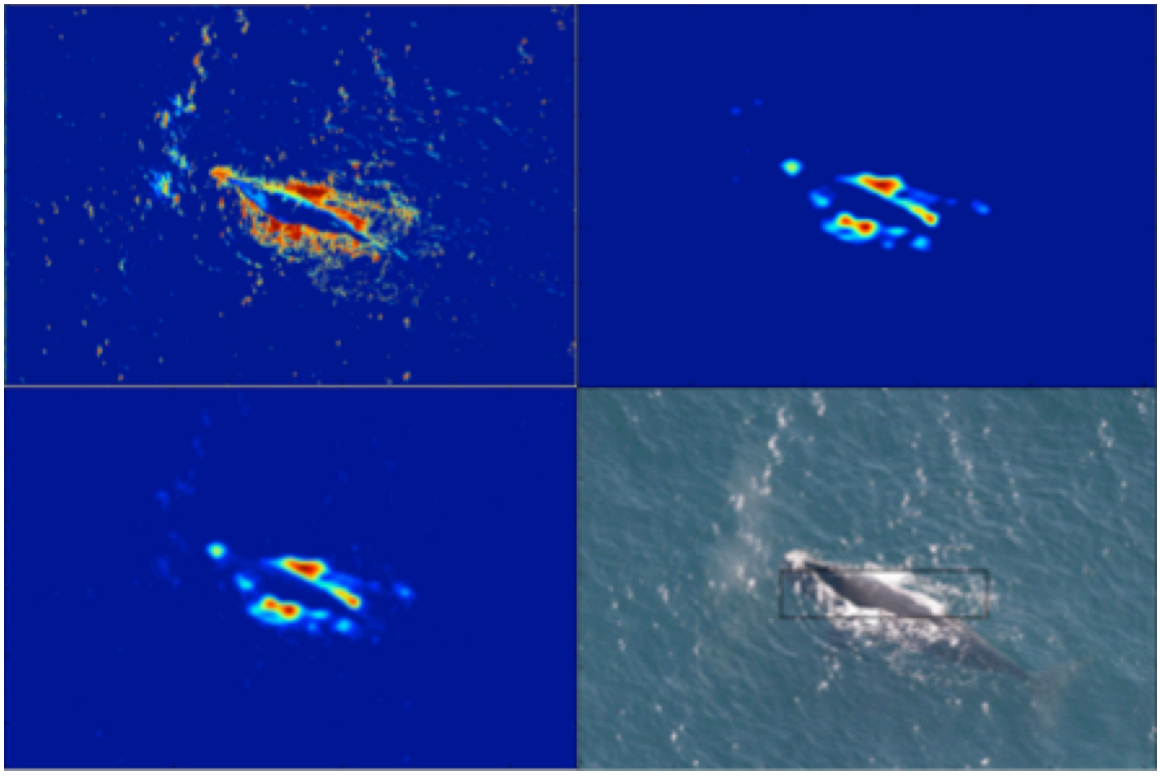

**Figure 5**: Localization failure. (Top-left) median subtraction and normalization, (Top-right) EMD distance matrix image. (Bottom-left) aggregated image. (Bottom-right) bounding box failure.

Figure 5 shows that waves are still being accounted for in median subtraction part of the algorithm, and histogram with Earth Mover's Distance is not a very good way to distinguish whale from the background when callosities are not visible and waves are around the whale. Those waves, since they were forming a boundary to whale's outlines, were considered to be the region of interest.

**Conclusion and Future Direction**

In the future, there should be a series of descriptors that account not only for the general local feature such as corners and gradients, but neural network of such features giving spatial context. Such feature space will capture what simple bag-of-words classifier could not, especially in the case where object's outline is not distinct and background noise is very similar in shape and color to object feature we are looking for.

*Code: https://github.com/jungsw/cs231a*

**References**

Ofir Pele and Michael Werman, "Fast and robust earth mover's distances," in Proc. 2009 IEEE 12th Int. Conf. on Computer Vision, Kyoto, Japan, 2009, pp. 460-467.