# Project: Right Whale Recognition Kaggle Competition

# by Sunday Srinivasan, Anton Rosenbloom and Eric Dybdahl.

The competition statement is straight forward: given 11468 JPG images of right whales with 4544 set aside as a training set can you come up with a machine learning algorithm that will predict which 447 whales match with the remaining 6924 images.

Right whale identification is performed in the field via visual inspection of the callosity pattern on the whale’s head. The goal was to automate the task of indentifying whales based upon the distinctive callosity patterns.

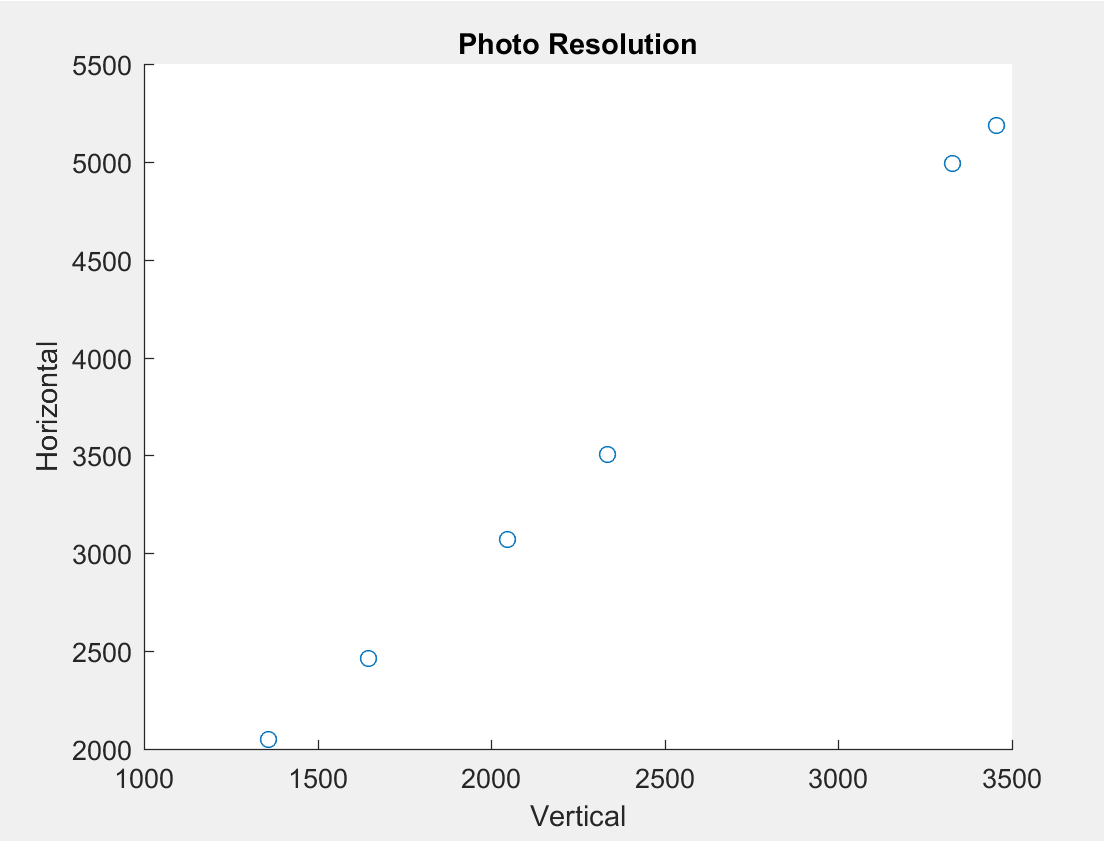
With this goal in mind, the task involved 3 sub-tasks: 1) Exploratory Analysis and Preprocessing of images (Eric), 2) Extracting features from images using the approach of Eigenfaces for recognition (Sundar) and 3) ML based classification (Anton).

These sub-tasks and lesson’s learned are detailed in the following.

# Data Analysis and image processing (Eric Dybdahl)

The Images are the dataset. There is no other attribute and the whale ids or targets are assigned to only the training set of 4544 photos. The image processing problem is to both modify and classify the photos adequately so that a recognizer can be applied to the photos.

The photos are all aerial shots taken from different angles under varied lighting and ocean conditions. The photo resolutions varied but aspect ratios remained constant as shown in the following scatter plot and histogram of the 4544:





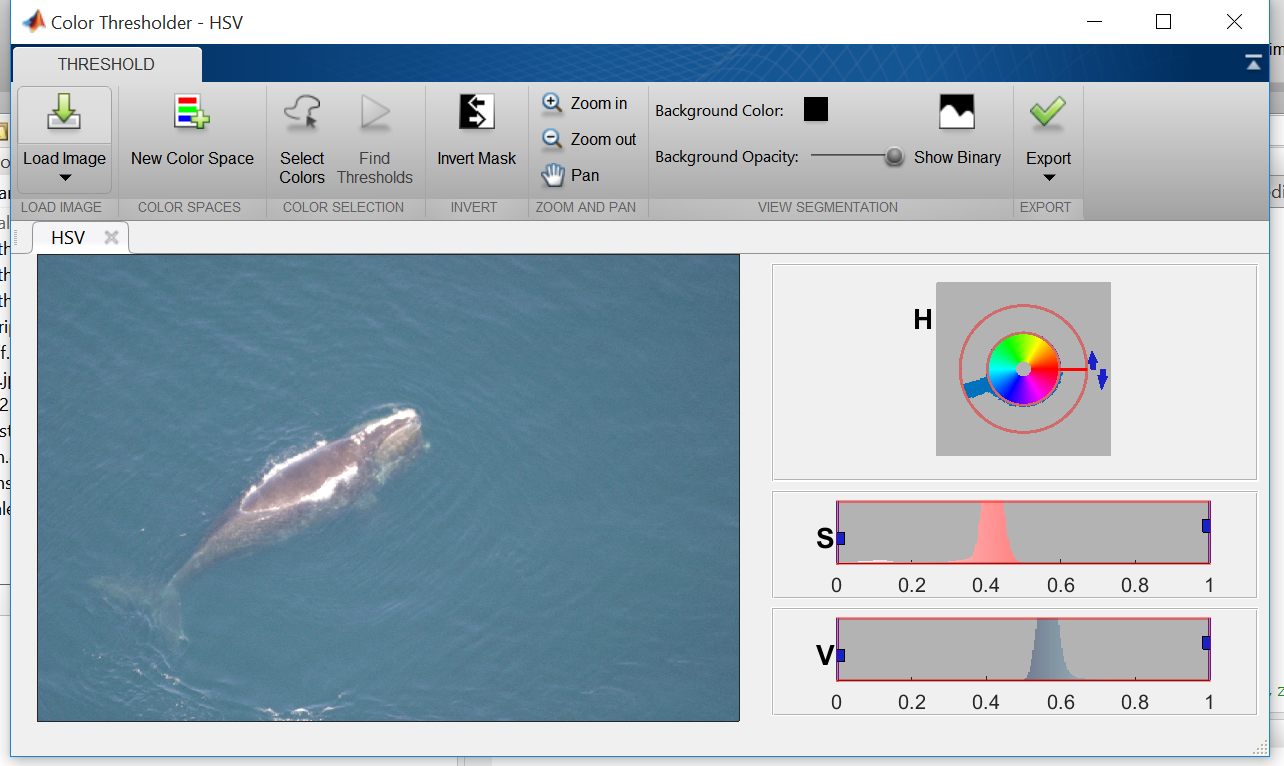
With the majority being 2048x3072.

We are trying to use a facial recognition algorithm on the whale images which means that we need to crop down to an image of the whale that is as uniform as possible. The image processing problem to be solved are as follow:

1. Orient the whale horizontally in the image.
2. Isolate the head and crop down to it.
3. Remove the obscuring spray from the whale image or find a set with minimal spray

The methods used in attempting to solve these problems were color separation, edge detection least squares fit and affine transformation.

We were able to achieve through the use of HSV color separation a distribution of the whale and a distribution of the spray for each photo. A typical HSV color distribution for a given whale is illustrated below:



Were there are five distribution regions for these photos:

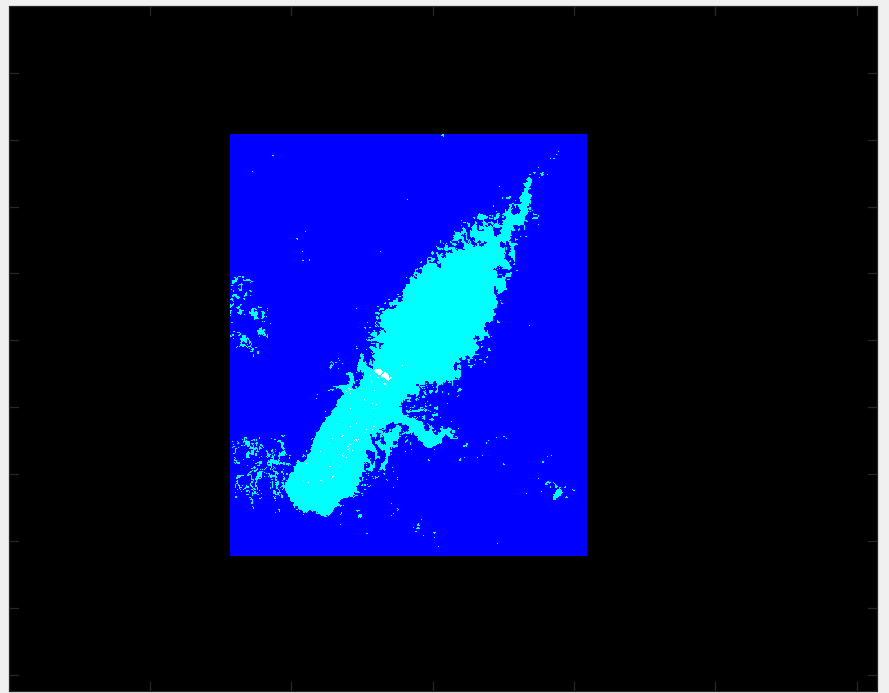
1. One in the H or hue range centered around 5.5 and describes mostly the blue green color of the water.
2. One in the S or saturation range above 0.2 and show strong in colorful photos.
3. One in the S range below 0.2 which defines most of the gray whale and the spray.
4. One in the V or value range below 0.8 which defines the most of photos light intensity.
5. One in the V range above 0.8 which defines the spray and the whales white head ornament.

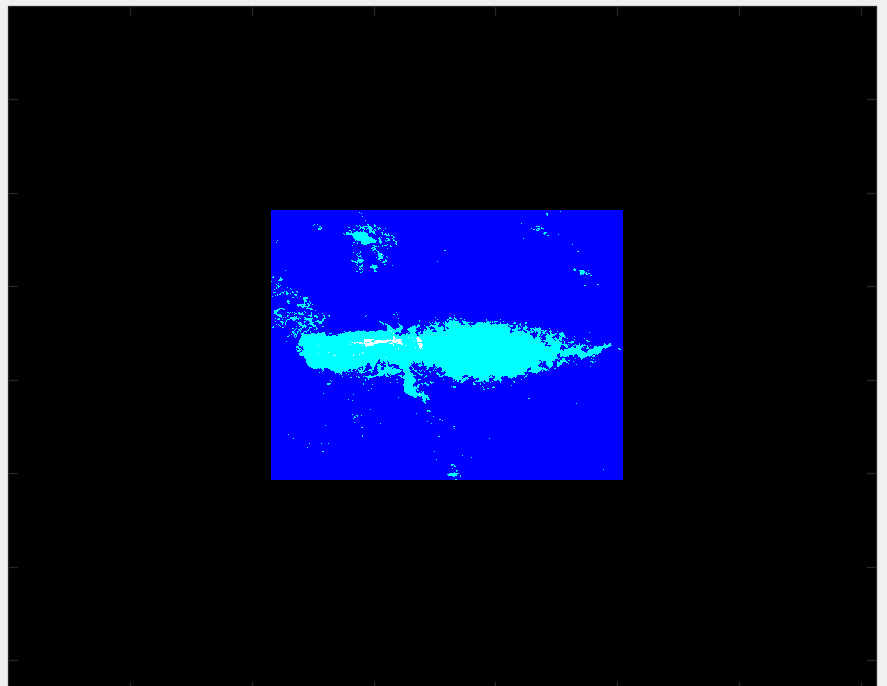
The intensity, mean and variance and existence or absence of these distribution characterize a whale photos color range.

The whale distribution can be pulled out with V < 0.8, and S < 0.2 or H > 0.6. The distribution of the spray with V > 0.8.

With the use of morphing techniques the whale distribution can be cleaned up so that we can get a center of mass and create an ROI, region of interest and zero out anything outside of it. Then run a least squares line through the center of mass. Using the center of mass and the slope of the line the photo can be translated and rotated with an affine transformation such the whale is both centered and horizontal. Cropping down on the whale distribution a new photo is created that is a centered and horizontal whale.









The direction of the head can be found by a color separation of the white head ordinate and the spray. It happens that in general the head ordinate is more yellow and slightly more color saturated than the spray. By calculating the center of mass for the head ordinate the direction of the head can be determined and rotated so that the head is facing right if necessary. Then the left half can be zeroed out so that a new center of mass and ROI can be calculated and used to crop down on the head and resize the image to 1024x1536.



With adjustments most of the whale images can be cropped down by using this method. Unfortunately only through visual examination of the results are we be able to know if the image was cropped down successfully. It would be nice to have some way of scorning the final images. Also this method failed on about 10% of the photos. The conclusion is that color separation needs to be done in concert with other image and object recognition and analysis methods.

As of yet we have not attempted any adjustment for the spray since we did not have a measure of how this would affect the recognition results.

This method was used to orient whales along a common spatial direction, crop into equi-sized images and constituted 1 set of image data that was then processed via the EigenFace algorithms so as to extract features for whale classification.

# Eigen-face based Whale Recognition (Sundar Srinivasan)

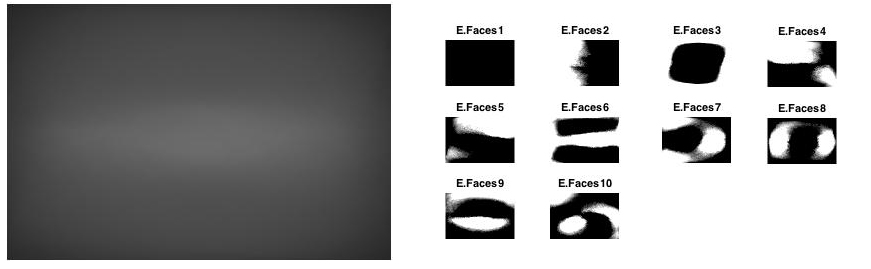
From the “T” training images (each RGB image is MxNx3 matrix of pixels), we determined whale ‘eigenfaces’ using the approach of Turk and Pentland (“Eigenfaces for Recognition”, J Cognitive Neuroscience, 3:71-86, 1991). Briefly, from the 4544 training images (cropped and oriented by Eric, previously), we determine eigenfaces and keep the top M as the eigenface basis that could explain 90% of image space variance.  We then projected each of our training set images (known ids) into the face basis and to obtain a vector of their ‘eigen weights’. These weights served as the set of ‘features’ that were then used to train models and classify images by whale ID.

Then for each new ‘test’ image, the image (transformed into a M\*N\*3 vector for RGB images; into a M\*N vector for Grayscale images) was projected into the whale eigenface basis to obtain the set of weights associated with the test image, and used in concert with the trained model to classify the whale ID associated with the test image.

In our first attempts at using this algorithm to classify images, we used the approach of Euclidean distance between the test eigen weights vs that of the training eigenweights to assign whale IDs to the test image. The idea was to ultimately use these weights to design more sophisticated, ensemble ML models for classification (by Anton).

However, some issues rapidly became apparent including memory limits that required conversion of RGB images to gray scale and resizing images to 25% of original size (which required acceptance of the accompanying loss of image resolution). This issue could be alleviated if the whole whale images could be further cropped to include just the head of the whale.

We then applied the Eigenface algorithm to the cropped oriented images that Eric had generated at a first pass, and images of the resulting ‘average’ whale (merging of all training images), and the top 10 whale eigenfaces are shown below:



The eigenface basis, and associated weights were then determined for each of 4544 training images. This data set was then partitioned into a 70:30 train-test split, Euclidean distance used to classify (the test images) and the accuracy of classification was determined.

For the 1st instance where whales were oriented and cropped, the classification accuracy of the eigenface based recognition method was extremely poor at 0.22 + 0.15%.

This result may not have been surprising, given that the number of pictures per unique whale was extremely skewed in the training set (ranging from n=1 to n=47 images/unique whale). Additionally, the first pass with orientation of whales and cropping had some misses (apparent on visual inspection, but not otherwise). The other possibilities for this failure included: 1) the noisiness of the images where whale’s images were obtained across a range of lighting conditions, camera angles whale positions (e.g., anterior-posterior to medial-lateral or face up – sideways – face down positions), 2) use of the whole whale image vs cropped to head (where the distinctive, identifying callosities are located), 3) the range of whale ‘pixel’ resolution (due to camera elevation), 4) loss of information in conversion from RGB to grayscale, 3) loss of image resolution due to memory issues, and 5) the inappropriateness of the Eigenface approach for the classification of noisy image data. We attempted to (somewhat) systematically address each of these issues by:

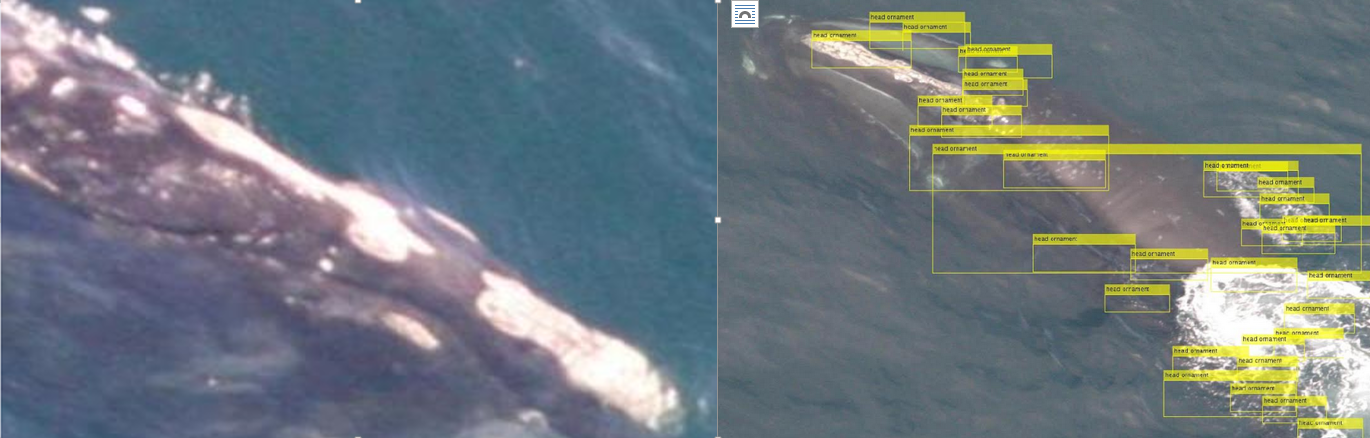
1. Researching alternative means to better orient, crop and possibly identify features that could be used to ID whales (Anton).
2. Grouping whales (exploratory clustering) so that the large training set could be split up into smaller sub-sets based on similarity measures
3. Exploring the appropriateness of the Eigenface approach for whale recognition (Sundar and Eric).

# Alternative Methods for Imaging Processing/Feature Extraction (Anton)

It’s become clear early on that to achieve a high rate of accuracy in recognition of individual whales we would have to identify and extract (reduce the original images to) well-defined Regions Of Interest (ROI). Furthermore, even to be able to make the educated decision as to which parts of the whale bodies can serve as identifying “markers “ in classification we need the ability to generate image-sets where the particular part of the body is isolated and oriented uniformly. While we were quite successful at isolating head-shots of the whales using color-based methods described above in the “Data Analysis and image processing” section, we recognized the need in more flexible approaches, the kind that can be easily tuned for discriminating for different types of RIOs.

Hence, the described below research in Object Detection, the end-goal of which is to be able to scan the raw images (utilizing the “sliding window” method) for target ROIs. This work is not completed. Results shown below are preliminary.

Our “first stab” at the task was to evaluate tools readily-available at our disposal and highly accessible (easy to use), such as cascadeObjectDetector tool within Matlab’s Computer VisionToolbox. The framework through its interactive tool *imcrop* allowed us to quickly create a set of 178 positive training images of the “head ornament” and train the cascadeObjectDetector model (based on Viola-Jones algorithm) using it and the programmatically created matching set of negative training images. This approach is widely used for person and face-detection. The representative result of this is shown on the figure below. Note, that while multiple areas in the image were identified as “head ornament”, all of them are along the whale’s body and none in the surrounding water.



"Head ornament" (left). Detection results (right)

CascadeObjectDetector framework is highly automated and largely functions as a black-box with internal image pre-processing and classification (it uses SVM). It is not clear, for example, how it handles the difference in training image sizes.

Next, we moved to test various classification algorithms for the task of Object Detection. For representation we chose Histogram Orientation Gradient features (HOG) widely used in Computer Vision for person and face detection in images. Below (left-to-right) are the original image, the image laid-over with its representations with HOG features (CellSize 16 and 32) and CellSize 32 HOG representation on its own.

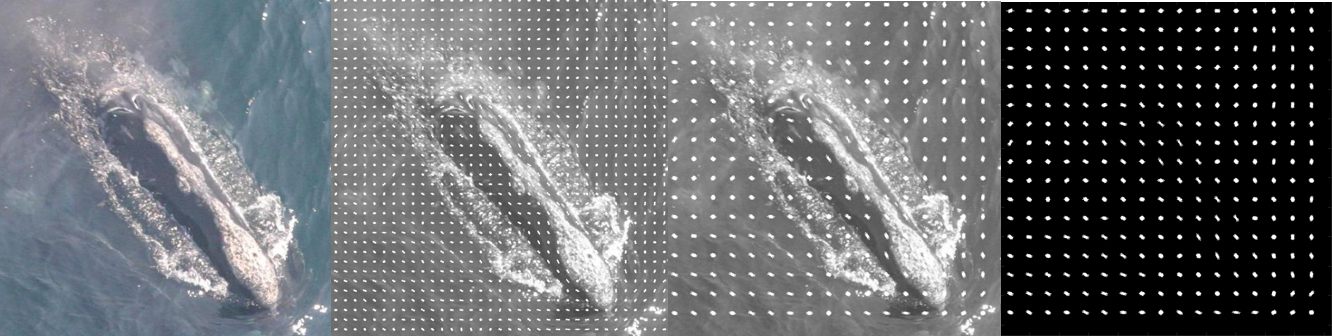


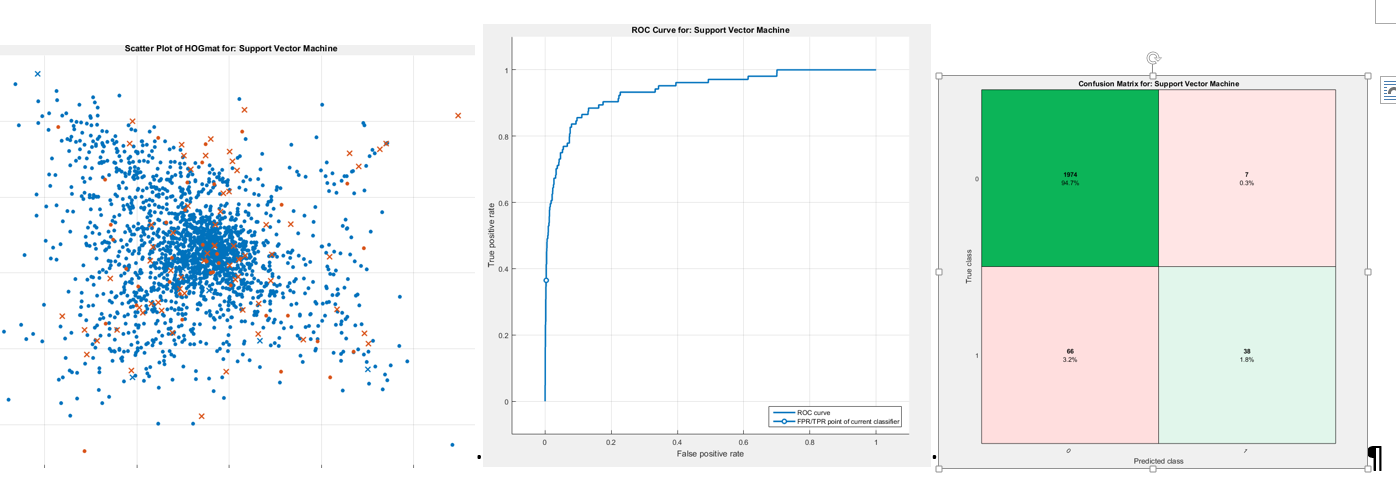
Image and its HOG-feature representations

The main attraction of the HOG representation is that it’s much sparser than the original image yet retains the key features of it.

Having learned from the previous experiment that the training sample set must be much larger and that the N of negative samples must be >> than the N of positive samples, we’ve generated a new training set of 104 positive samples of head-shots and 2000 negative samples.

We faced a choice of how to deal with difference in sizes of the training images when creating the training matrix for the classification algorithms: the default parameters of the exractHOGFeatures function resulted in feature-vectors of vastly varying size for different images. We chose to re-calculate the CellSize parameter for each training image so that the feature-vector for each image would be of the same specified length (not exact, but very close) regardless of the image size.

We’ve tested different classification models (SVM: linear, quadratic, cubic), decision trees, and KNN on training matrixes of length of 20000, 2000, and 1000 features – with PCA and without. Below shown the best training result to date – a cubic SVM trained with 5-fold cross-validation on a 1000-feature long training matrix of 2100 some training samples with no PCA reduction. Overall accuracy of classification (in cross-validation) 96.5% with the ROC curve of 94%. Given the heavy skew of the training set towards the negative samples, the accuracy figure is somewhat misleading. While correctly identifying almost 100% of the negative samples, it misclassifies 63% of the positive samples (FNR=63%).



Scatter plot (left). Note that some brown dots are classified correctly; ROC curve (middle); Confusion Matrix (right)

This concludes the research into Object Detection methods performed to date. Admittedly it is barely “scratching the surface” of the subject.

# Exploratory Clustering of Images (Eric Dybdahl)

In part due to the (poor) classification accuracy The 4544 test images are very diverse in brightness and color saturation. In order to classify the images better HSV color distribution statistics for the five distributions were added to the train table. As follows:

hsum, hmean, hvar: is the magnitude, mean and variance for the hue distribution.

vlsum, vlmean, vlvar: is the magnitude, mean and variance for the lower value distribution

vhsum, vhmean, vhvar: is the magnitude, mean and variance for the high value distribution.

slsum, slmean, slvar: is the magnitude, mean and variance for the low saturation distribution

shum, shmean, shvar is the magnitude, mean and variance for the high saturation distribution.

Preformed unsupervised learning kmean clustering on the following columns, to establish a classification of photos: hmeam, hvar, vlmeam, vhmean, slmean, shmean. K of 4 was the most stable and the centroids are as follows:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Size | hmean | hvar | vlmean | vhmean | slmean | shmean |
| Cluster1 | 1501 | 0.5644 | 0.0034 | 0.5090 | 0.9241 | 0.1028 | 0.4368 |
| Cluster2 | 912 | 0.5540 | 0.0105 | 0.4757 | 0.9149 | 0.1138 | 0.2694 |
| Cluster3 | 1280 | 0.5552 | 0.0054 | 0.3538 | 0.9239 | 0.0933 | 0.5015 |
| Cluster4 | 798 | 0.5117 | 0.0185 | 0.2994 | 0.8955 | 0.1038 | 0.2788 |
| NaN | 53 |  |  |  |  |  |  |

Cluster 1 – brightest photographs from vlmean=0.5090 and high color saturation of shmeam=0.4368. Water color is in a consistent blue-green color range; hmean=0.5644 with a narrow variance of hvar=0.0034.



Cluster 2 – still bright but grayer photographs with vlmean=0.4757and low color saturation of shmeam=0.2694. Water color is more variable between a grayer green to blue range; hmean=0.5540 with a larger variance of hvar=0.0105.



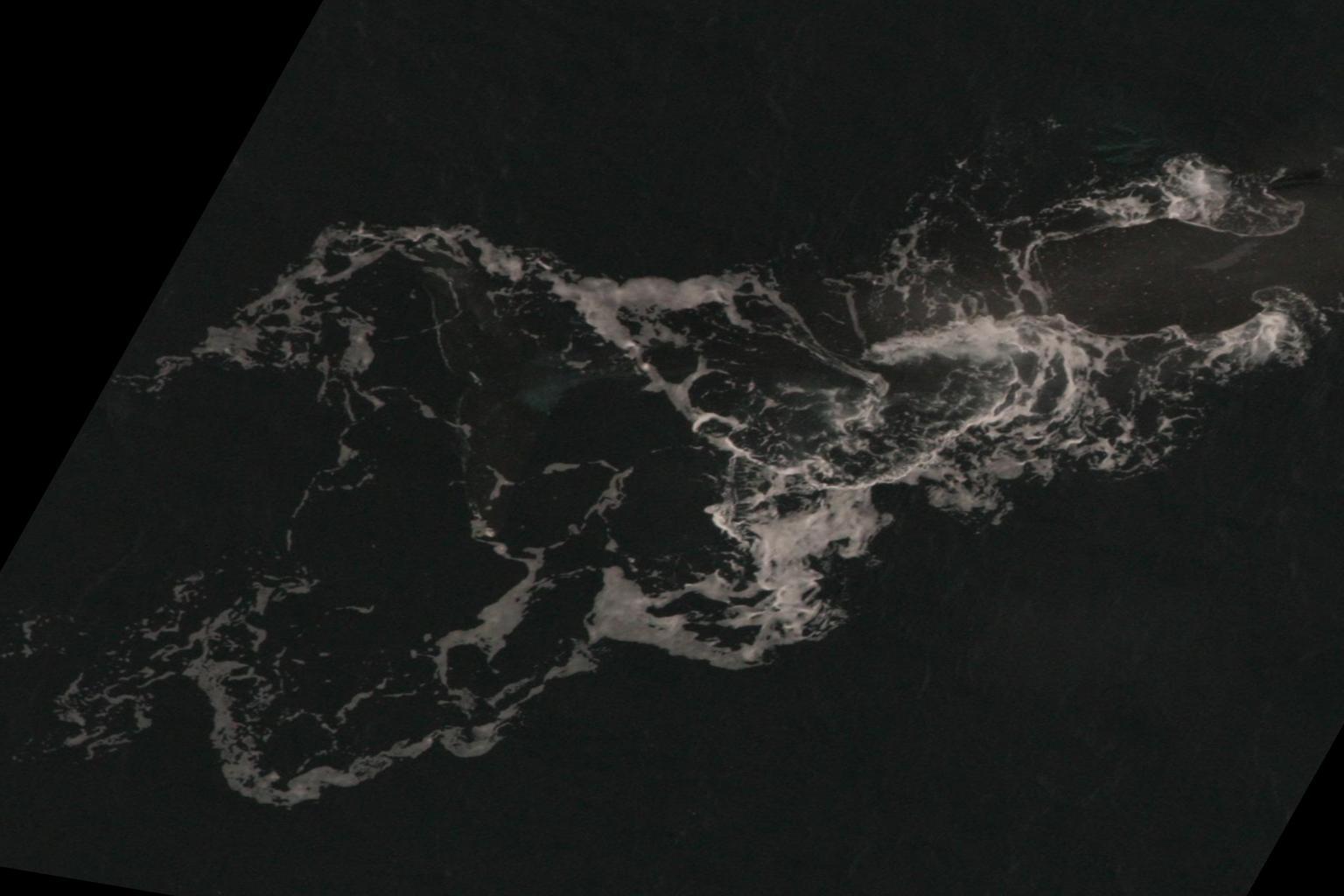
Cluster 3 – are darker but still have high color saturation with vlmean=0.3538 and shmean=0.5015.The water Color is in the narrow blue-green range: hmean=0.2994 with a narrow variance of hvar=0.0054.



Cluster 4 – are both bark and gray with vlmean=0.2994 and shmean=0.2788. The water color is mostly gray with hues ranging form green to blue: hmean=0.5117 with a larger variance of hvar=0.0185.



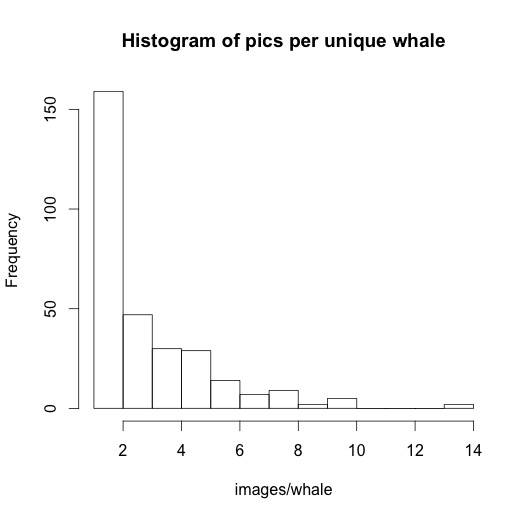
NaN – which could be considered cluster 5 have no distributions for either high value or high saturation. They are either very black or very gray.



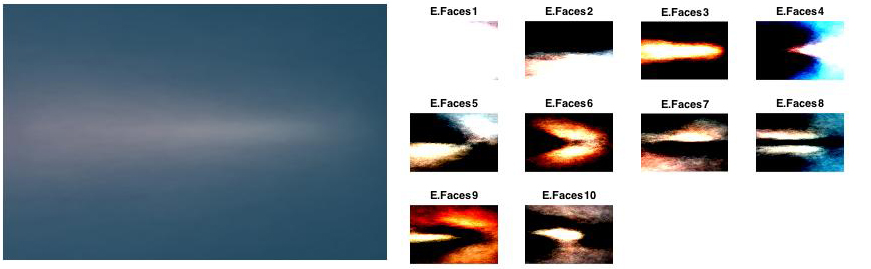


# Appropriateness of Eigenface Approach (Sundar and Eric).

As part of determining the appropriateness of the Eigenface approach, we restricted analysis to a set of images identified previously to be in Cluster 1. Furthermore, these images were processed and cropped to only include the head of the whale. This subset (cluster 1) constituted 929 images of 304 unique whales, but with a skewed histogram of images/unique whale (a problem with train-test randomization for classification)!



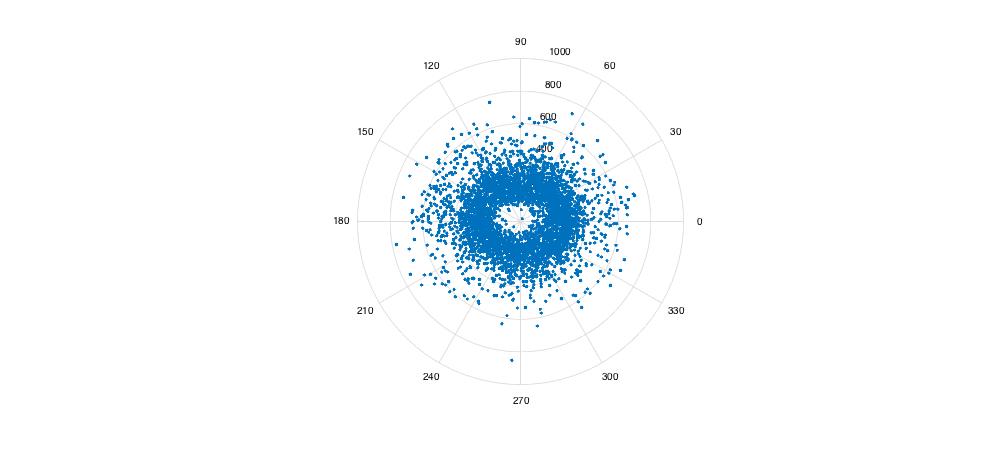
For this subset (n=929), we utilized the Eigenface – Euclidean distance based classification to explore the classification accuracy as a function of whether the eigenfaces were extracted from: 1) color images at 25% of original resolution, 2) Gray scale images at 50% resolution, 3) Grayscale at 25% resolution, 4) Black and White images (binary) thresholded at a pixel value of 100 (in a 0-255 range), 5) BW images at a threshold of 150, 6) BW images at a threshold of 200, 5) BW images at a threshold of 225. A sample image of the average whale and top 10 eigenfaces from the RGB images of this subset is illustrated below:



The classification accuracy for these different experimental conditions are plotted in fig above. We found that the accuracy across these settings improved from 0.22% (for whole whale) to ~2% (Fig below), and image resolution/content did not significantly influence classification accuracy (i.e., RGB, grayscale, BW at different thresholds). As such, while extracting features from images of whole whales oriented in similar directions to images of whale heads appeared to improve classification accuracy, the overall accuracy remained extremely poor (~2%).

The possibilities underlying this improved, but still poor result could be: 1) observations that whale head cropping/orientation was not error free, 2) the negative impact upon accuracy from randomization of train-test split, given the skew in images/whale, 3) the inappropriateness of the eigenface based feature extraction for classification of noisy data (i.e., images of whales obtained over a 30-yr period in a variety of settings/resolutions).

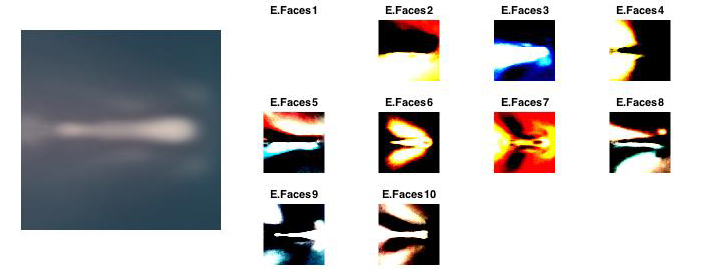
In part to address issues with orientation/cropping errors, we accessed a publically available data set (by Anil Thomas, a Right Whale Kaggle competitor) on the spatial coordinates of the blow holes and bonnet tips of all whales in the training data set. Given these points, we were able to determine the size of heads (as distance between blow hole and bonnet tip) as well as the orientation of the whales in the original images (See below).



This analysis suggests that the images capture whales oriented over 360 degrees, but also, that the ‘head’ size ranges from 19 to 850 pixels, i.e. a 44-fold scale difference. Assuming that differences in adult whale sizes, particularly head size cannot span such a large range, the conclusion was that the images of the whales were obtained at a widely ranging resolution.

We therefore used the data (on bonnet tip, blowhole positions) to process images such that the whales are oriented in the same direction (due East), cropped to head (using the bonnet tip/blowhole positions post re-orientation) and resized such that whale heads across images had the same length (as the average). These images were then used in a classification exercise (80:20 random spit) using the eigenface feature extraction and Euclidean distance based classification scheme.

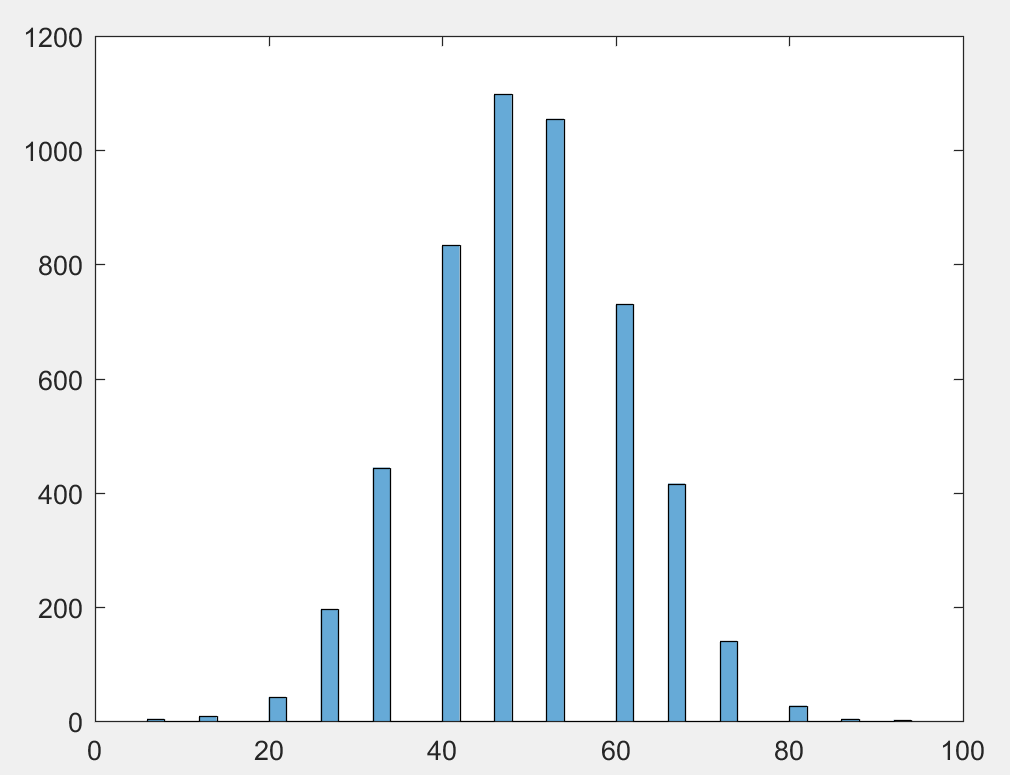
A sample image of the average whale and top 10 eigenfaces from the RGB images of this subset is illustrated below:



Visual comparison of the average whale with our previous two attempts (using the entire data set of images oriented, but not cropped and oriented/cropped) suggests that the third approach yields the best set of whale head images for further analysis. And correspondingly, the classification accuracy also improved to 10 + 2% from 2.2%, but nevertheless remained poor and suggested the need for clustering/further image processing and further analysis of the appropriateness of the eigenface based approach for feature extraction.

# Analysis of the Eigenface recognizer on a subset of the best cropped photos. (Eric Dybdahl)

So far we lacked any good measure of what constitutes a well aligned clean image for a recognizer like Eigenface. In order to find that I removed some of the obstacles in process that were getting in the way of that evaluation. 1. By creating script so that I could frame the head accurately by selecting two points on the image manually. 2. Restricting to set of photos to six whales that have at least 10 photos in the cluster 1, 78 photos. Taking 5000 runs using a 80:20 training to test split produce this distribution of accuracy:



The mean of the distribution is about 48%. Which for six whale is better than randomly assigning a photo to one of six whales which would be a distribution with a mean of about 1/6 ~ 17%. Even though the Eigenface recognizer may not be the best for this messy data set it can can still be used to measure improvements in image processing like alignment color correction or reduction or the smoothing of spray. My confusion is that the Eigenface recognizer is very sedative and though I may be a poor choose as a general recognizer for this problem it can be used to measure improvements in image processing and given better scoring of the images it could be used as concert with other recognizers to vote for the best match.

# Conclusion

The central issues that to this problem has been determining how to best process the photos. From the investigation of the photos and the process there are clearly three parts to this problem that need to be solved in an iterative fashion with proper feedback:

1. Imogene processing (how clean I,e. Allined, color corrected or enhance the photos need to be for recognition.
2. A way to measure or score the appropriateness or class of the photos for recognition.
3. An image recognizer or set of image recognizers to be used in the evaluation.

What I have seen as a solution to similar problems is a pipeline process where each step refines the image for the next step in the process and each step requiring a separate skill. The messiness of the photos has been stated throughout this paper. It has been demonstrated that color separation can be used to isolate the head but that alone cannot determine if the resulting cropping contains the head or how well it is framed. However, it can give another process a high probability of where to look. With a ROI defined by color separation we can use an object recognizer like HOG and/or image templates and convolution to refine the ROI and give a score to the image existence, possible orientation, resolution and class. With this information through experimentation and training an appropriate image recognizer can be chosen and/or vote for the appropriate recognizer.

Our first run, using the Eigenface recognizer with attempts at using color separation for whale head and alignment resulted in 0.22 + 0.15% accuracy, showed us that we were lacking impotent steps in this process. We have been spending the rest of our time reevaluating the process to find measures and techniques to improve it and evaluating how far we really were from a real pass at a ML model.

The two paths we took were to evaluate the appropriate the Eigenface recognizer and to find better techniques in determining ROI, object recognition and image classification. The conclusion are as follows:

1. The Eigenface recognizer yet powerful is far too sedative to be used on the photos as we can get them processed currently. We ether need to look for other less sensitive recognizers or enhance photo quality.
2. Color separation alone is not adequate enough to process and score these photo. We need to look at other techniques like object recognition, HOG and/or image templates with convolution in a pipeline alone with color separation and possibly photo enhancement techniques.

We are still in the process of evaluating new techniques as you can see from the rest of the paper. We have not run out of problems or question.