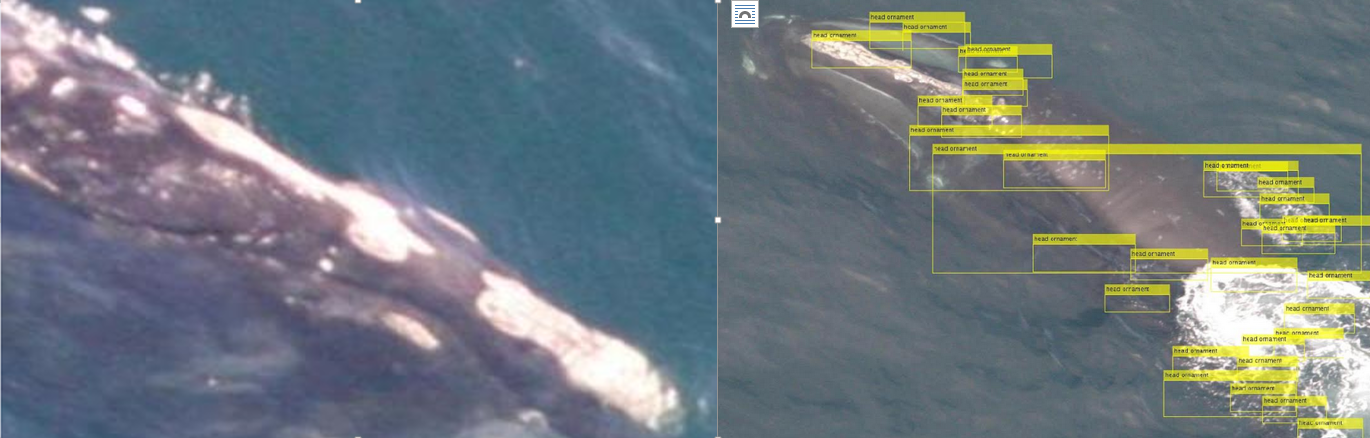
# Alternative Methods for Imaging Processing/Feature Extraction (Anton)

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.

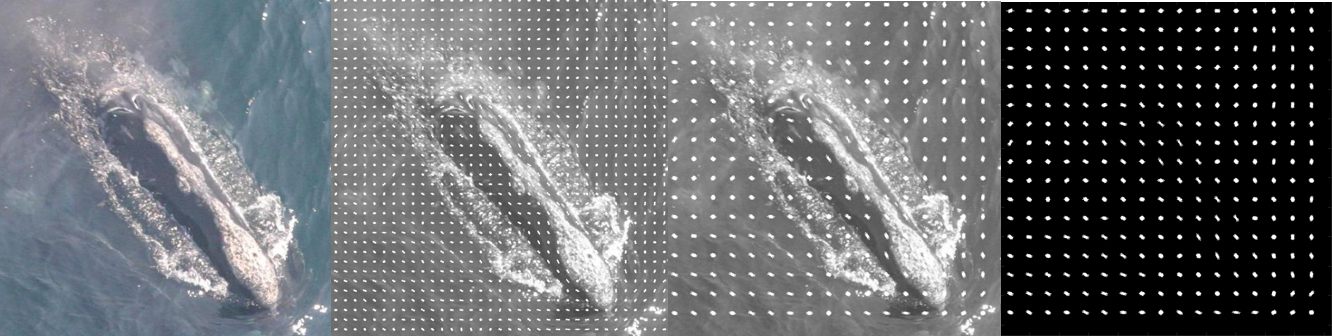


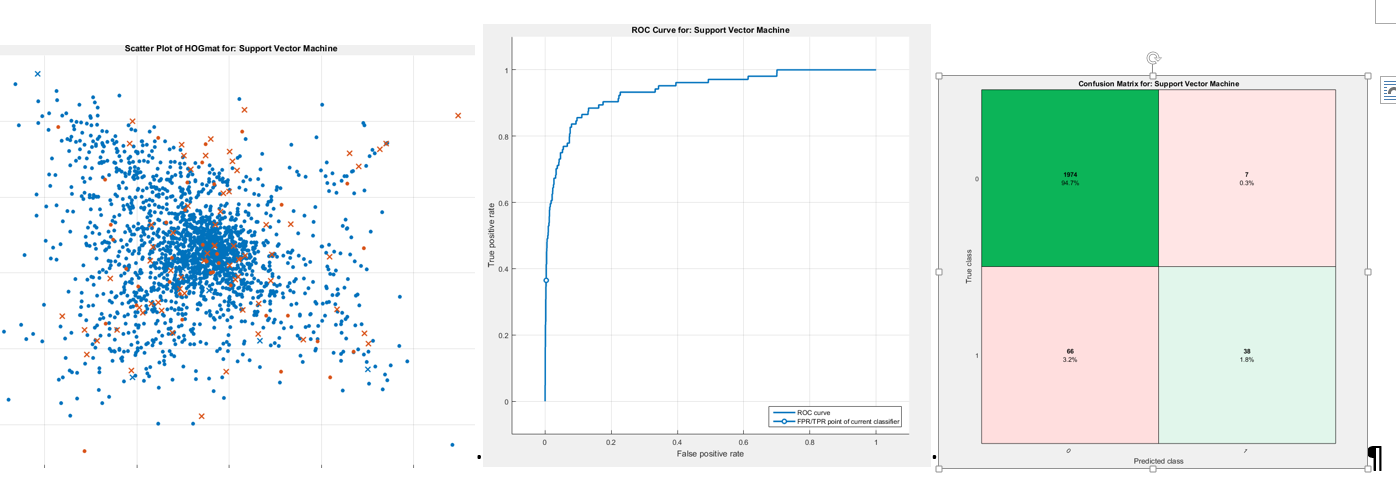
Image and its HOG-feature representations

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.