# Sonar Independent ATR

Oliver Daniell
Heriot Watt University
SeeByte Ltd., Edinburgh, UK
Email: oliver.daniell@seebyte.com

Yvan Petillot
Ocean Systems Lab
Heriot Watt University
Edinburgh, UK
Email: yvan.petillot@seebyte.com

Scott Reed
SeeByte Ltd., Edinburgh, UK
Email: scott.reed@seebyte.com

***Abstract*—This paper presents a model-based, filter response algorithm for Automatic Target Recognition (ATR) in Sidescan Sonar (SSS). The filters are created from the projection of a large set of generating boxes. The first and second-order statistics over the highlight and shadow regions of the projected boxes form a feature vector for each pixel in the image. The algorithm selects the features which best describe the target object, during training. When the algorithm is applied to an image, the projection of the generating boxes under rotation and translation is used to approximate invariant regions of the object under the equivalent transformation. The performance of the algorithm is compared to the standard Haar cascade. It is shown that the algorithm presented in this paper has a reduced dependence on the image formation model, requires a lower number of features to train and matches the performance of the Haar cascade. Operationally this brings two key advantages. A single classifier can be trained on data from several different models of sonar without a significant loss in performance. Additionally, the algorithm uses a smaller feature vector, reducing the time required to train the algorithm from several days to under one hour. This makes it possible to train the algorithm in-situ, increasing robustness with respect to new sea-floor types and environments.**

## I. Introduction

Sidescan Sonar (SSS) is used for surveying large areas of the sea floor for mine-like targets. In current manual operations the data is searched visually by an operator. This is time consuming and subject to human error [1]. Automatic Target Recognition (ATR) algorithms can assist operators to identify mine-like targets. This reduces the time taken to analyse a large survey and increases confidence in the results. As the performance of ATR systems increases they are also being used for decision making on autonomous vehicles. These applications require algorithms to have a high Probability of Detection (PD), a low Probability of False Alarm (PFA) and to run efficiently on embedded systems.

Supervised ATR algorithms typically have two stages. In the detection stage potential objects are identified in the image. In the classification stage features are extracted, from a window about the target, and a decision about the object class is made by a machine learning algorithm. This approach works very well when the training data is similar to the testing data. However, it can produce poor results due to changes in the environmental and operational conditions. Bell et. al. [2] propose two approaches to this problem. The first approach introduces independence to variations in the image formation using a Model-based algorithm [3]. Here, a target prior is used to model the dependence of the ATR on the image formation model; the model is then matched to contours extracted from objects in the image. The algorithm is very sensitive to the initial segmentation of the image. For this reason it is unable to match the performance of algorithms which extract features directly from the image. Conversely, algorithms which extract features directly from the image are more sensitive to changes in the image formation. Therefore, they are less able to generalise to changes in the environment. Augmented reality training images, where simulated targets are inserted into real background images, are proposed for training these algorithms in new environments. For this to be practical the algorithm must be suitable for training in-situ as part of a survey mission.

The current state of the art for ATR in sonar image uses the second of these two approaches [4]. The algorithm comprises a boosted cascade of decision trees with a Haar feature set [5]. The key difference between this algorithm and previous filter response algorithms [6]–[8] is that the feature set is over-complete. The classifier selects the most relevant filter response from the feature set during training. Therefore, while any single feature is insufficient to describe the target, the ensemble of features creates a very strong classifier. The Haar cascade has excellent run-time performance, however the large number of features makes it very slow to train on current hardware. Therefore the Haar cascade is not suitable for in-situ training.

In this paper a combination of these two approaches is proposed. An image formation model is used to define locally adaptive templates. The templates are used to measure statistical properties directly from the image. Typically model-based algorithms require a different model for each target. To avoid this, the algorithm approximates the object using a plurality of generating boxes of different dimensions. The highlight and shadow regions of the generating boxes are projected into the sonar image and statistical properties of the image measured over these regions. Rotation and translation of the target is therefore modelled by an equivalent transformation of the generating boxes. Additionally, changes in the speed of the vehicle or the range of the sonar are corrected for using the local image formation model.

The algorithm is tested against the standard Haar-cascade. It is shown that comparable performance is achieved, over the same data set, with a considerably smaller feature vector. This reduces the training time from over a day for the Haar cascade to under 30 minutes. Therefore, it it possible to train the algorithm in-situ, to cope with changes in the sea-floor environment. The robustness of the image formation model is demonstrated by training the algorithm on data collected from one model of sonar and applying it to a different sonar type without retraining. This approach allows Mine Counter-

978-0-933957-40-4 ©2013 MTS

Measure (MCM) operations to be conducted with new vehicles and sensors without the initial cost of building a new target database. Additionally the results can be incrementally improved by retraining the algorithm in-situ with new background data collected from a previously unseen region.

## II. Description of Algorithm

In this section the two stages of the algorithm are discussed. Sections II-A and II-B cover the initial image processing and generation of skewed integral and squared skewed integral images. These "helper" images allow the large number of statistical calculations to be performed in real time. This process is outline in figure 1.

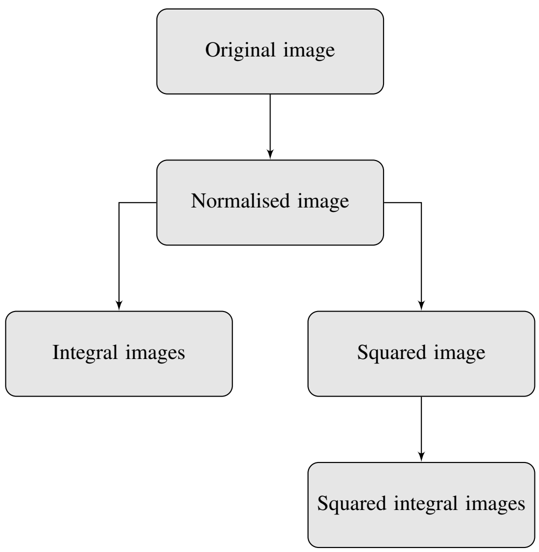


Fig. 1. Integral images are calculated from the normalised image and the square of the normalised image. These "helper" images are used to accelerate the calculation of the mean and variance of regions in the normalised image

Sections II-C to II-D cover the derivation of the feature set and the classification algorithm. A set of generating blocks are projected into the sonar image using ray tracing [9]. First and second-order statistics are calculated over the highlight and shadow regions of these templates. There are 8 features associated with each block 5 first order statistical features and 3 second-order statistical features. Modelling the change in appearance of the projection of the block over the image reduces the dependence of the 8 associated features on the image formation model. Individually these features are weak. Therefore, an ensemble of features is used to create a robust description of an object. A cascade of boosted decision trees (Section II-D) is used for feature selection and classification [10].

### A. Image Normalisation

The aim of image normalisation is to reconstruct the image that would have been produced independent of factors such as beam pattern and sonar loss [11]. Features can be normalised using the local background intensity, however this prevents the absolute value of the pixels from being used as a feature. This results in false alarms in dark regions of the image where the noise is amplified. Simple approaches which estimate a single intensity curve for the entire image [12] can result in object sized artefacts. The method of Dobeck [13], using adaptive filters, tracks paths of similar intensity through the image. The filters follow surface returns and beam patterns effectively removing unwanted artefacts from the image. The set of integer skewed integral images is calculated from the original image normalised using this method.

### B. Integer Skewed Integral Images

Integer skewed integral images are images in which subsequent rows of the image have been skewed by an angle $arctan(\alpha/\beta)$ where $\alpha$ and $\beta$ are integer constants. The integer skewed integral images approximate the appearance of the projection of the generating blocks in SSS. The formulation is similar to that of the standard integral image with the addition of a skew function $s(y, y')$ to calculate the horizontal shift between each row of the image. The integral image $I(x, y)$ is calculated from the normalised image $i(x, y)$ as

$$I(x,y) = \sum_{\substack{y' \leq y \\ x' \leq \max(0, x - s(y,y'))}} i(x', y') \tag{1}$$

where

$$s(y, y') = \left( \left\lfloor \frac{y}{\beta} \right\rfloor - \left\lfloor \frac{y'}{\beta} \right\rfloor \right) \times \alpha \tag{2}$$

A number of skewed integral images, corresponding to different skew angles, are calculated. The regions that can be calculated from these images are constrained by the skew angle of the image. The skewed rectangular region, defined in figure 2 can be calculated as

$$\sum i(x', y') = I(A) + I(C) - I(B) - I(D) \tag{3}$$

where

$$\begin{gathered} A(y) < y' \leq C(y) \\ A(x) - s(y', A(y)) < x' \leq C(x) - s(y', A(y)) \end{gathered} \tag{4}$$

And valid positions for A,B,C,D are

$$\begin{aligned} A(y) &= B(y) \\ C(y) &= D(y) \\ A(x) &= D(x) - s(D(y), A(y)) \\ B(x) &= C(x) - s(B(y), C(y)) \end{aligned} \tag{5}$$

The skewed integral images are used to approximate the highlight shadow geometry produced by a box located on the sea floor. By varying $\alpha$ and $\beta$ as in table I an approximately uniform distribution of box angles relative to the sonar array can be created.

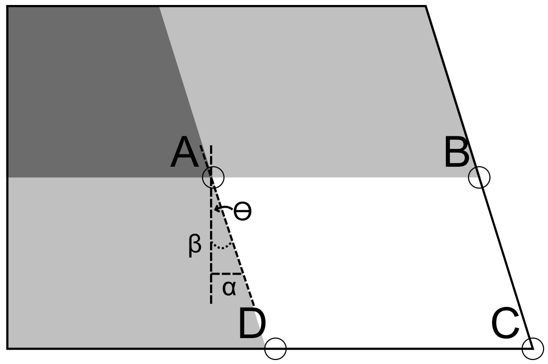


Fig. 2. An integral of a region can be reduced to a function of four points in the integer skewed integral image representation

| $\alpha$ | $\beta$ | Angle |
|---|---|---|
| -4 | 1 | -76° |
| -2 | 1 | -63° |
| -1 | 1 | -45° |
| -1 | 2 | -27° |
| -1 | 4 | -14° |
| 0 | 1 | 0° and 90° |
| 1 | 4 | 14° |
| 1 | 2 | 27° |
| 1 | 1 | 45° |
| 2 | 1 | 63° |
| 4 | 1 | 76° |

TABLE I. AN APPROXIMATELY UNIFORM DISTRIBUTION OF SKEW ANGLES BETWEEN $\pm 90°$

### *C. Normalised 3D Features*

The features are defined using an image formation model. A set of generating blocks with across track dimension $x$, along track dimension $y$ and height $z$ are used to create the features. For each generating box, first and second-order statistics are calculated over the regions defined by the projection of the box into the SSS image. The projection is calculated locally for each pixel, given the grazing angle $\phi$ and the local image resolution. The sonar image formation is approximated using ray tracing as shown in Figure 3. The size of the highlight region in the across track direction $h_x$ is related to the grazing angle $\phi$ as $h_x = x/\cos(\phi)$. The shadow length $s_x$ is $s_x = z/sin(\phi)$. The orientation of the target is not known a priori. The projection is calculated for the rotation angles defined in table I. A rotation $\theta$ of the target is approximated as a skew of the highlight and shadow regions. This allows the features to be calculated efficiently using the integer skewed integral images. For an angle $\theta$ the highlight size in the across track direction $h_x = \frac{x}{\cos(\phi)\cos(\theta)}$, the shadow size $s_x$ is invariant under rotation. The Highlight and shadow size in the along track direction varies with $\theta$ as $h_y = s_y = y\cos(\theta)$.

A number of features are defined for each generating block, 5 first-order features and, 3 second-order features. For each feature the most target like value over all rotations of the generating block is selected as the feature value. The features are described in table II and the highlight and shadow regions shown for a number of rotation angles in figure 4. The first feature defined in table II is the mean of the pixel values in the highlight region. For each generating block there will be 11 feature values corresponding to 11 rotations of the block. The rotation angle which best matches the rotation angle of the target will contain only the backscatter from the target. The maximum value from the 11 rotated values is selected as the highlight region is brighter than the surrounding sea-floor and object shadow. Therefore, the 11 values can be reduced to as single feature which is independent of rotation. A similar logic can be applied to reduce the other 4 first-order features from 11 values to a single value. The second-order features calculate the variance over the highlight and shadow regions. When these regions match the rotation of the object it is assumed that the variance will reach a local minimum.

| Name | Description | Selection rule |
|---|---|---|
| Highlight | The mean of the highlight region | Maximum value |
| Shadow | The mean of the shadow region | Minimum value |
| Object Template | Shadow subtracted from highlight | Maximum value |
| Bounded Front | Object template with additional rectangle bounding the front of the object (fig 4) | Maximum Value |
| Bounded Side | Object template with additional rectangle bounding the side of the object (fig 4) | Maximum value |
| Highlight Var | Variance of the highlight region | Minimum value |
| Shadow Var | Variance of the shadow region | Minimum value |
| Object Var | Highlight Var + Shadow Var | Minimum value |

TABLE II. DESCRIPTION OF THE 8 FEATURES CALCULATED FOR EACH BLOCK. THE MOST TARGET LIKE VALUE OVER ALL ROTATIONS IS SELECTED USING THE SELECTION RULE

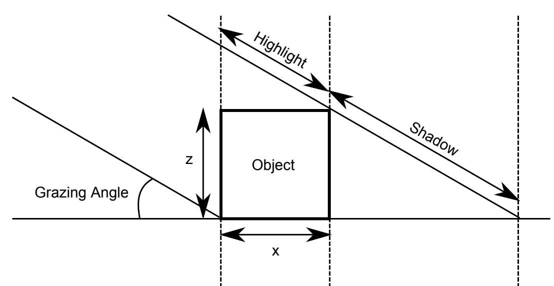


Fig. 3. A model for the size of the highlight and shadow regions for a cube like object. The object has height z and width x and the Highlight and Shadow length are calculated for an image which has not been slant angle corrected.

### *D. Feature Selection and Classification*

The feature set is large and it would be intractable to calculate all features for every pixel. Therefore, boosted decision trees are used for feature selection. Multiple boosted decision tree classifiers are combined using a cascade structure. Boosted decision tree cascades use the concept of weak and strong classifiers. Weak classifier are those for which the output is only weakly correlated to the true class (i.e. the classifier is

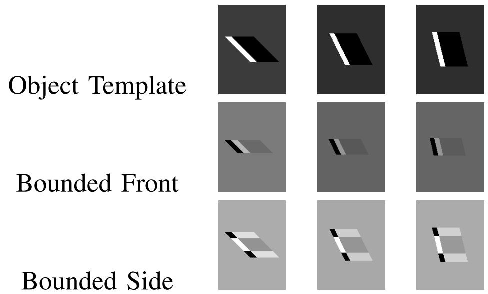


Fig. 4. Example of the extended feature set. The Bounded Front and Bounded Side features can be used to eliminate objects which are larger than the target.

slightly better than a random guess). Strong classifiers can be arbitrarily well correlated to the true class.

A strong classifier can be created from a set of weak classifiers through boosting. This is shown in figure 5. The weighting of the training samples is altered such that samples classified correctly receive a lower weighting when evaluating the error for the next weak classifier. The strong classifier is created by summing the output of the weak classifiers multiplied by a weighting coefficient $\alpha_t$. This weights the classifiers according to their error over the training set. Using stump decision trees, consisting of a single node, each weak classifier is essentially a cut over a single feature. Thus, the boosting algorithm selects the most informative features as a consequence of building the strong classifier.

The majority of the background in SSS images consists of flat featureless sea-floor. Evaluating an image on a pixel by pixel basis is costly, therefore to minimise the number of features evaluated for simple background regions a cascade of classifiers is used. A cascade is a sequence of strong classifier where each classifier in the sequence is trained and evaluated on the regions that are evaluated as target like by the previous classifier. By tuning each strong classifier such that it retains greater than 98% of the positive samples and rejects 50% of the background samples a large proportion of the image is rejected with relatively few calculations. This process is demonstrated in figure 5.

Each pixel in the image is evaluated by the classifier. A confidence value is obtained by weighting the output of the boosted decision tree stages of the cascade, following the approach described in [14]. Each stage $n$ produces a confidence value $c_n$. The final confidence value $C$ applies a greater weighting to those stages that occur later in the cascade.

$$C = \sum_n n * c_n \tag{6}$$

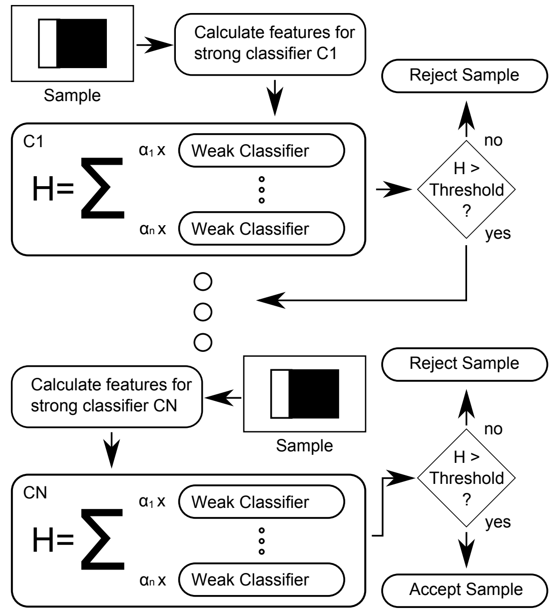


Fig. 5. A cascade of boosted weak classifiers

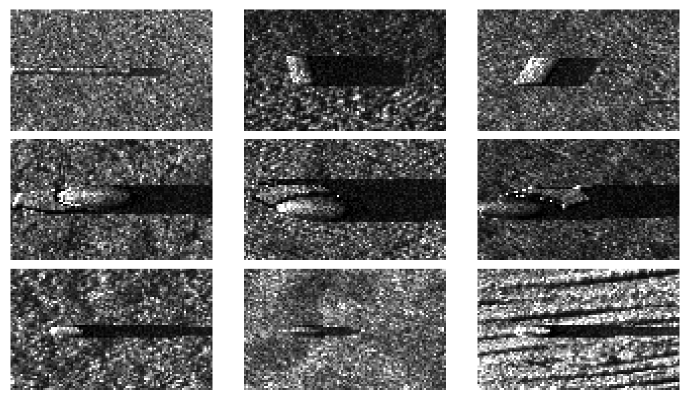

Fig. 6. Examples of the augmented reality targets. Cylinder (Top) Complex (Middle) Truncated Cone (Bottom)

## III. Description of Data Set

The data set consists of 2000 Edgetech and MarineSonics images randomly sampled from various tow-fish and Autonomous Underwater Vehicle (AUV) missions. These images include flat, rippled and cluttered regions of sea-floor. The images are randomly split into a training and a testing set and augmented reality objects inserted into the image [15]. The target set consists of a long narrow cylinder, a complex asymmetric object made up of several small spheres and a symmetric truncated cone (Figure 6). The algorithm is trained on a data set consisting of positive target regions and background images which may contain other target types.

## IV. Results and Analysis

The algorithm is trained and tested on the targets shown in figure 6. The performance characteristics for the three targets are shown for the Edgetech data set in figure 7. The truncated cone and the complex classifier both achieve a PD better than 95% for less than 20 false alarms per square kilometre. The cylinder classifier achieves a lower maximum PD and an order of magnitude more false alarms. The PD can be attributed to the low along track resolution of the Edgetech sonar and the asymmetry of the cylinder target. Cylinders with their major axis oriented in the across track direction have an along track signature which is only 1 pixel wide (Figure 6 top left). The low number of pixels increases the probability that random noise and clutter will have a target like signature. The large number of false alarms can be attributed to several rippled areas which are visually identical to the end on aspect of the cylinder. In cluttered and rippled regions which are not identical to the targets the false alarm rate is similar to that of the other two targets.

The results for the selectivity of the classifiers are shown in table III. The results are taken at the point in the Receiver Operator Characteristics (ROC) curve where the curve intercepts with the line at 90 degrees to the no-discrimination line (Youden's J statistic). The results cannot be used to produce a confusion matrix as each classifier is a binary classifier. Some objects in the test images will be classified as target by all three classifier, to produce a final classification for these objects an additional decision stage would be needed. The classifier specified in the row heading is applied to the target type specified in the column heading. The cell contains the percentage of targets classified by the classifier. The results demonstrate that less than 2% of real targets will be incorrectly classified by the algorithm.

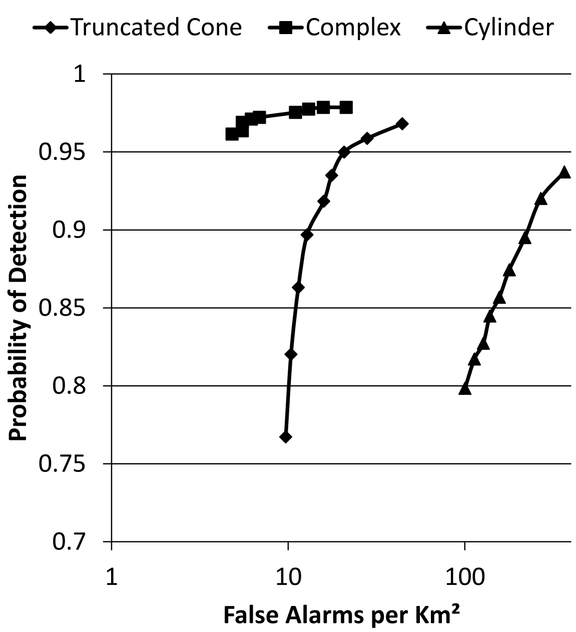


Fig. 7. Performance characteristics for the classifier trained and tested on Edgetech data.

The cylindrical target is used to test multi-sonar classification. Applying the algorithm without modification produces poor results due to the algorithm over specialising on the statistical information specific to each sonar. Applying a median filter to the images removes much of this variation. It can be seen in figure 8 that the cylinder classifier trained on Edgetech data performs well on the MarineSonics data.

While it has been shown that reducing the statistical variation between images can produce a more generic classifier, we want to retain this information to obtain a high PD. Therefore, where we have training data for all the desired sonar types the algorithm can be trained on a combined data set. Figure 9 shows the performance characteristics for a cylinder classifier which has been trained on both Edgetech and Marinesonics data. The performance is better than that for the classifier trained without the statistical data, however it should be noted that the algorithm still achieves a PD of more than 80% even on unseen sonar data.

The performance of the algorithm is compared to the standard Haar cascade on the Marinesonics data set in figure 10. Both algorithms were trained with 16 cascade stages. Each stage was trained to accept 99% of the target objects and reject 50% of the background data. Both algorithms perform well with a high PD and low PFA. The Haar cascade demonstrates approximately half the number of false alarms for any given PD compared to the algorithm presented in the paper. Both algorithms detect the majority of cylinders with the major axis parallel to the direction of travel but are unable to detect the end on cylinders. This is potentially due to the low resolution of the sonar and the low contrast of the target in the end on aspect. The algorithm presented in this paper is more robust

| Classifier | Object Type | | |
|---|---|---|---|
| | Truncated Cone | Cylinder | Complex |
| Truncated Cone | 95% | 8% | 4% |
| Cylinder | 1% | 91% | 0% |
| Complex | 1% | 3% | 98% |

TABLE III. RESULTS FOR EACH CLASSIFIER TESTED ON THE OTHER TARGET DATA SETS. ERROR BARS ARE CONTAINED WITHIN THE POINTS

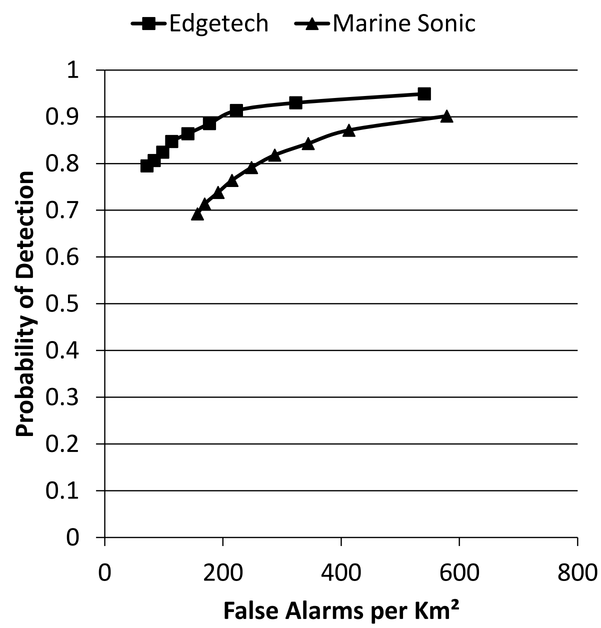


Fig. 8. Performance characteristics for a cylinder classifier trained entirely on Edgetech data and tested on MarineSonics data. Error bars are contained within the points

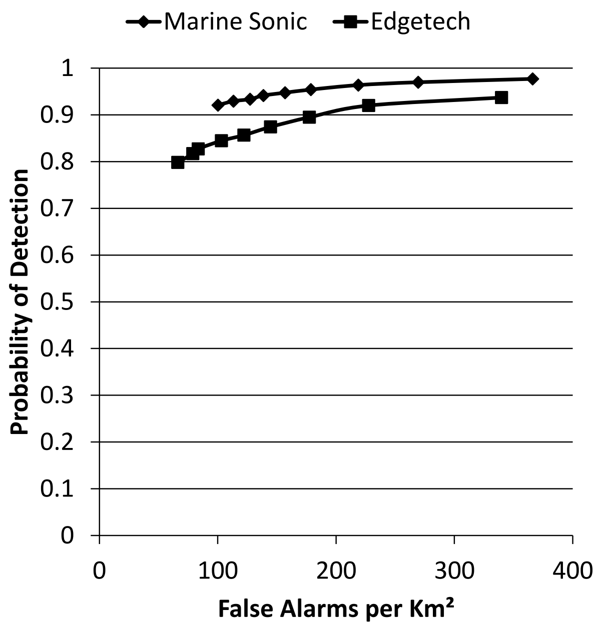


Fig. 9. Performance characteristics for the classifier trained and tested on MarineSonics and Edgetech cylinder data. Error bars are contained within the points

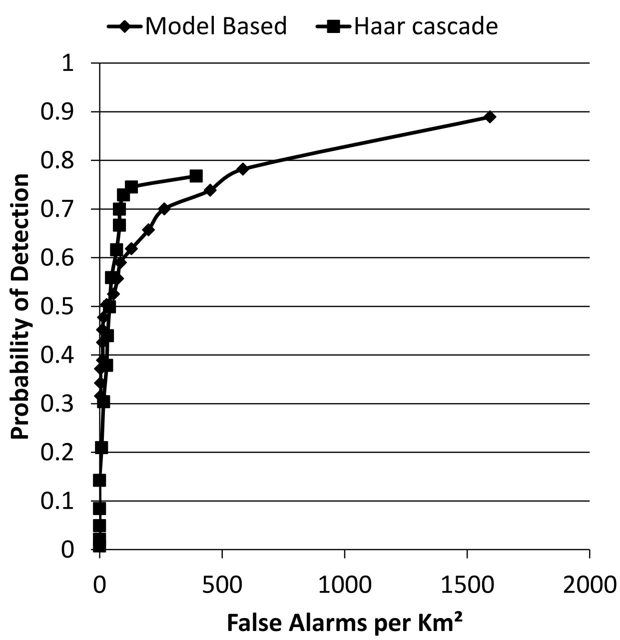


Fig. 10. Comparison of the performance of the standard Haar cascade compared to the performance of our algorithm. Error bars are contained within the points

with respect to rotation and detects an additional 10% of the testing data set.

## V. Conclusion

This paper has demonstrated a novel algorithm for mine-like object detection in SSS. The algorithm combines the advantages of model-based and filter response algorithms to achieve a high PD, low PFA and excellent selectivity. It has been shown that the algorithm is relatively independent of the model of sonar used. The algorithm detects 80% of targets on a completely unseen sonar type with only 400 false alarms per kilometre squared. Surprisingly, training the algorithm on a combined data set of Edgetech and Marinsonics data improves the performance over both testing data sets. In this case it is possible that the larger training data set increase the ability of the classifier to reject previously unseen background images.

The algorithm demonstrates approximately double the number of false alarms for a given PD when compared to the standard Haar Cascade. However the maximum PD achieved by the algorithm is 10% higher than that of the Haar cascade. The algorithm presented in this paper takes approximately 30 minutes to train, compared to over a day for the Haar cascade. This makes it possible to train the algorithm on the local environmental conditions, in-situ, as part of an operation. It will be the subject of future work to test if in-situ training negates the superior PFA demonstrated by the Haar cascade.

An interesting consequence of this approach is that the feature space is only weakly dependent on the orientation or grazing angle of objects on the sea floor. We intend to adapt the geometry of the locally adaptive features such that they can be applied to forward looking sonar as well as SSS. In theory this could allow vehicles with different sensor payloads to operate using the same feature space and share a common picture of the environment. Potentially this could enable a vehicle to localise itself in a SSS map using the data observed by the forward looking sonar.

## Acknowledgment


The authors would like to thank the EPSRC for funding this research




## References


[1] P. Chapple, "Automated Detection and Classification in High-resolution Sonar Imagery for Autonomous Underwater Vehicle Operations." in *Proc. Spie*, 2008. [Online]. Available: http://dspace.dsto.defence.gov.au/dspace/handle/1947/9897

[2] J. Bell, Y. Petillot, S. Reed, E. Coiras, P. Y. Mignotte, and H. Rohou, "Target Recognition in Synthetic Aperture and High Resolution Sidescan Sonar," in *IET Proc. High Resolution Imaging and Target Classification*, 2006, pp. 99–106.

[3] S. Reed, R. I. Tena, C. Capus, and Y. Petillot, "The fusion of large scale classified side-scan sonar image mosaics." *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 15, no. 7, pp. 2049–60, Jul. 2006. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16830923

[4] N. V. Y. Petillot, Y. Pailhas, J. Sawas and J. Bell, "Target Recognition in Synthetic Aperture Sonar and High Resolution Side Scan Sonar Using AUVs," in *International Conference on Synthetic Aperture Sonar and Synthetic Aperture Radar*, 2010.

[5] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, May 2004.

[6] E. Coiras, P. Mignotte, and Y. Petillot, "Supervised target detection and classification by training on augmented reality data," *Radar Sonar Navigation IET*, pp. 83–90, 2007. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4119406

[7] M. Granara, A. Pescetto, F. Repetto, G. Tacconi, and A. Trucco, "Statistical and neural techniques to buried object detection and classification," in *MTS/IEEE Proc. Oceans Conf.*, vol. 3, 1998, pp. 1269–1273.

[8] Q. Huang, M. Azimi-Sadjadi, B. Tian, and G. Dobeck, "Underwater target classification using wavelet packets and neural networks," in *IEEE Trans. Neural Networks*, vol. 1, 1998, pp. 177–182.

[9] J. M. Bell, "A Model for the Simulation of Siclescan Sonar," in *MTS/IEEE Proc. Oceans Conf.*, no. September, 1995.

[10] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, pp. I–511–I–518, 2001.

[11] C. G. Capus, A. C. Banks, E. Coiras, I. Tena Ruiz, C. J. Smith, and Y. R. Petillot, "Data correction for visualisation and classification of sidescan SONAR imagery," in *Proc. Spie*, vol. 2, no. 3, 2008, pp. 155 – 169.

[12] S. Anstee, "Removal of range-dependent artifacts from sidescan sonar imagery," *DSTO*, 2001. [Online]. Available: http://dspace.dsto.defence.gov.au/dspace/handle/1947/3648

[13] G. J. Dobeck, "Image normalization using the serpentine forward-backward filter: application to high-resolution sonar imagery and its impact on mine detection and classification," in *Proc. Spie*, vol. 5794, no. 381, 2005.

[14] J. Sawas, Y. Petillot, and Y. Pailhas, "Cascade of Boosted Classifiers for Rapid Detection of Underwater Objects," in *Proceedings of the European Conference on Underwater Acoustics*, 2010.

[15] P. Y. Mignotte, J. Vazquez, J. Wood, and S. Reed, "PATT : A Performance Analysis and Training Tool for the Assessment and Adaptive Planning of Mine Counter Measure ( MCM ) Operations," no. Mcm, 2009.