ADAPTIVE AUTOMATED DETECTION FOR SYNTHETIC APERTURE SONAR IMAGES USING SEABED CLASSIFICATION

JA Fawcett DRDC Atlantic Research Centre, Dartmouth, Nova Scotia, Canada
WA Connors DRDC Atlantic Research Centre, Dartmouth, Nova Scotia, Canada

1 INTRODUCTION

It is well known that the performance of sidescan and synthetic aperture sonar automatic target recognition (ATR) methods depends significantly upon the seabed environment[1-4]. In particular, ATR methods will typically have a high probability of detection with a low false alarm rate for benign or featureless seabeds, however, the same methods will often suffer a considerably higher false alarm rate for rippled or cluttered seabeds. Most detectors utilize a threshold on a filter or classifier output where these thresholds are typically selected by an expert for a specific data set. For a given ATR method, the threshold required for a specified detection rate or false alarm rate may vary considerably with respect to the environment. Furthermore, one may also wish to use different methods or training sets for different environments. For all these situations, an estimate of the seabed's parameters is required in order to adapt the ATR. Williams and Fakiris[2] have described the use of multiple classifiers, each trained using data from different seabed types and the automatic weightings of these classifiers' outputs using the local seabed parameters. Gazagnaire et al[4] also describe the use of multiple classifiers and the use of adaptive thresholds to improve performance.

The approach in this work is somewhat similar. The basic ATR method that is considered is the Haar-cascade[5,6]. This method is trained to discriminate snippets of target data from snippets of non-target (background) data. It seems somewhat intuitive that a method which is trained to discriminate against a certain seabed type will be the best detector when the same type of environment is encountered again. However, a trained cascade may still work well in another type of environment. In this work, a training set of sonar images will be clustered into 4 classes based upon the values of 2 features described later. Although it may be possible to obtain better clustering using more features, using two has the advantage that the distribution of feature values is easily visualized. Haar-cascade detectors and our implementation of them are briefly described. Five cascades are trained, a cascade for each of the background classes, and one cascade trained using all the backgrounds. A distinct set of data, but from the same trial as the training data, is then used for performance evaluation. The performances of some of the cascades with respect to the seabed parameters will be investigated with respect to probability of detection, false alarm rate, and threshold required for a specified false alarm rate. In addition, the empirical data will be used to train a model which can provide a smoother, continuous prediction of performance as a function of seabed features. Finally, a different set of trial data is used to investigate the robustness and generalizability of the detectors and their predicted performance. The data used in this paper was kindly provided by the NATO Centre for Maritime Research and Experimentation (CMRE), La Spezia, Italy.

2 SEABED IMAGE FEATURES

There have been many approaches for seabed classification or segmentation in the literature. Here, we follow the approach taken in Refs.7,8 and utilize only 2 features which try to differentiate rippled seabeds, cluttered and featureless seabeds. Our features are based upon the ripple-detector described in Ref.9. Let us denote a two-dimensional image matrix as I. The derivative of I in the row direction is denoted as I_x and in the column direction as I_y. Then the complex gradient image is formed as

$$I_\nabla = I_x + \mathrm{i}\, I_y \tag{1}$$

The square of the complex gradient has the form $I_\nabla^2 = r^2(x,y)e^{2i\varphi}$ where φ is the angle of the orientation of the gradient vector. This quantity will have relatively large values when averaged over a window containing imagery with a consistent gradient orientation. We will replace the derivatives of the image by the outputs of filtering with Haar rectangles of a set of different dimensions. These rectangles consist of the entries ±1 and are oriented horizontally or vertically. These filters can be quickly correlated with the image using the method of integral images[5,6]. Instead of the gradient vector of Eq.(1) we will consider

$$I_H(n_x,n_y) = I * V(n_x,n_y) + iI * H(n_x,n_y) \tag{2}$$

where V and H denote the vertical and horizontal filters and n_x and n_y denote the dimensions.
In the case of a periodic seabed, with a sufficiently large amplitude, there will be certain values of (n_x,n_y) for which $F(n_x,n_y) = \sqrt{|\,<I_H^2(n_x,n_y)>\,|}$ will be large. One of our features for a snippet of data is f_2 = mean(F) (with respect to the set of scales). A large value is indicative of a significantly rippled seabed. However, there could be cluttered seabeds for which this value could also be relatively large. There could also be seabeds with a periodic structure of small amplitude for which f_2 is not large. Another feature that is constructed from the $F(n_x,n_y)$ is that we consider the first and last rows and compute their mean values. The larger of these 2 values is then used and is denoted as μ_C. We then consider the feature $f_1 = \max(F)/\mu_C$ which is a measure of whether there is a preferred spatial dimension in F. In Figure 1, four representative image snippets and their corresponding Haar-filter outputs $F(n_x,n_y)$ are shown.

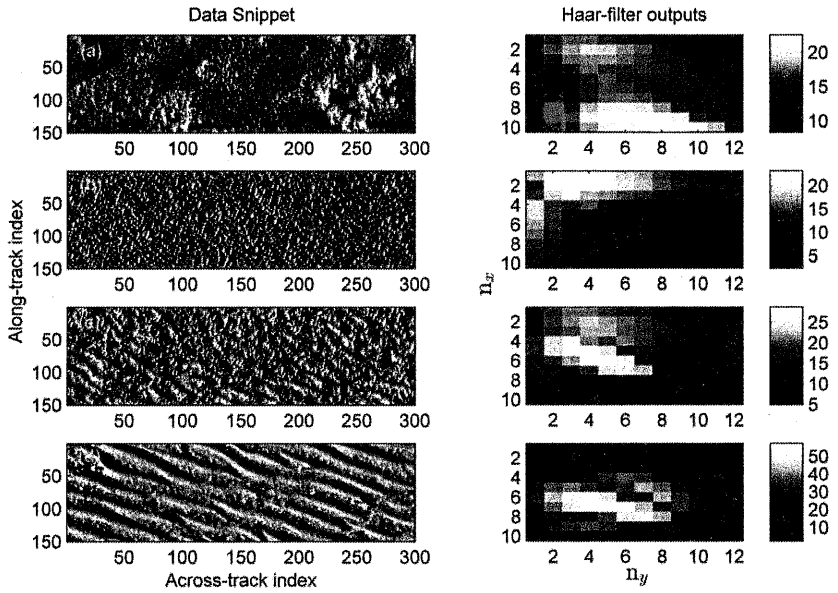

Figure 1 Four different data snippets and their F values

The two feature values corresponding to the snippets of Fig.1 are: (a) (1.22,16.02) (b) (1.38,10.67) (c) (1.79,15.45) (d) (3.34,22.59). The cluttered snippet (a) has a moderate value for feature 1 but a fairly low value for feature 2 while the snippet with small ripples (c) has a moderate value of feature 1 and a higher value for feature 2.

3 TRAINED PREDICTION MODEL FOR PARAMETER SELECTION

Below, we consider the average performance of a detector in discrete cells in seabed feature space. However, these cells are relatively large in size. Furthermore, the amount of data within each cell may vary significantly and also have a high variance. Building a regression model provides a continuous method for determining ATR parameters in-situ based on environmental measures. The model training was based upon the individual file data rather than on binned or cell-averaged results.

Support vector regression (SVR) is a kernel regression model using a support vector machine implementation. SVRs attempt to determine a linear regression function in high dimensional space through the use of a non-linear mapping function. The method was originally proposed by Vapnik et al.[10] to provide a method to determine a prediction function, $f(x)$, to describe the training examples provided to the algorithm, within a specific margin, or insensitivity tube, specified by the parameter ε. This insensitivity tube can be seen as the error margin of the function, as any training examples within this distance from the function will not impact the function chosen.

For the current work, a ε-SVR based on the libsvm[11] implementation was used to determine a function to describe the relationship between the environmental features sensed, the false alarm rate, and the threshold to be applied to the cascade detections. This model provides a method to adaptively tune the cascade based ATR in order to achieve a desired level of performance. This method provides the advantage of a prediction model for unseen environments, where cascade parameters can be predicted and the ATR system tuned in-situ. Furthermore, using fast training algorithms, the model may be retrained for future use when novel environmental feature values are encountered.

4 DATA PREPARATION

For training and validation, the CMRE Colossus II trial data was used. This data was obtained using the CMRE MUSCLE Autonomous Underwater Vehicle (AUV) equipped with a Synthetic Aperture Sonar. In this trial, 3 sites seeded with deployed dummy targets were surveyed. The seabed types varied at each of the 3 sites. The sites contained examples of seabeds with ripples, clutter, scours, as well as relatively featureless seabeds. In general, a single sonar tile may contain regions of different seabed types. The original sonar tiles had an along-track sampling of 2.5 cm and an across-track sampling of 1.5 cm. These files were pre-processed to remove large-scale amplitude variations, scaled and converted into TIFF files. For speed of processing, the images were decimated by a factor of 2 in the along-track direction and 5 in the across-track direction.

For this work, the sonar tiles, clutter files and files with targets, were divided into 2 sets – images alternately going into the training and validation sets. The training of the Haar cascades consisted of extracting small cropped images of the targets to form a set of "positive" images. The set size was increased tenfold by considering variations of the basic mugshot - the coordinates of the extraction window were randomly perturbed, the images were flipped vertically and resized versions were used. The directory of training clutter files provided the "background" files for the cascade training. It should be noted that for the MUSCLE sonar tiles there is a small amount of overlap with respect to the proceeding and following tiles, so that there are some small portions of the validation imagery which are also in the training set. To provide a previously unseen data set, images from the 2013 CMRE MANEX trial (from one of the sites) were used. During the trial, a number of standard dummy minelike objects were deployed and imaged with the MUSCLE SAS. The seabed in this area included regions of benign seabed but also more complex regions of Posidonia and clutter. In particular, Posidonia was a type of seabed not encountered in the COLOSSUS set. The TIFF images were prepared in an identical fashion as described above.

5 RESULTS

The training set clutter tiles are classified as one of 4 seabed types. The method for this classification is as follows. Each tile is divided into quadrants and will be subsequently saved as 4 separate images. To increase the number of samples available, we also considered vertically-flipped versions of the images. Within a single sonar tile, there can be a variation of seabed types, therefore by dividing the original tile into 4, some of this variation could be accounted for. For each image quarter, 7 random sub-windows of size 150 (along-track) X 300 (across-track) are extracted. Over these windows, the 2 features described in the previous section are computed. The two-dimensional distribution of these feature-values is then divided into 4 classes. This seemed a reasonable number of classes, however, more classes could be defined as required. We used a MATLAB Gaussian-mixture approach to determine these classes. The training set of clutter images were written into 4 different directories, corresponding to the most probable seabed class. (based on the mean probabilities from the Gaussian-mixture). The distribution of the features is shown in Fig.2. The Gaussian-mixture fit to the feature space is not unique and different clusterings are possible. Two such clusterings are shown. We used the top clustering for this paper. It should also be noted that the sampling of the feature space is very non-uniform. Due to the nature of the trial and the targets' seabed location, there is considerably more sampling of some sections of feature space. In this case, the more benign seabeds corresponding to lower values of f_1 and f_2 were sampled much more often.

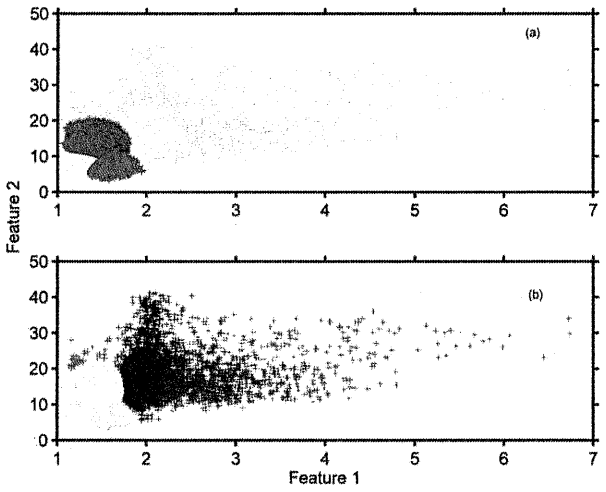

Figure 2 The distribution of features for the training set and 2 clusterings.

Four Haar cascades were then trained using OpenCV[12] with both positive images and clutter. An additional fifth cascade is trained using all the clutter images. In each case, all the positive images are used in the training regardless of the surrounding seabed type. The trained cascades are then used with images from the validation set using a multi-scale face detection algorithm from OpenCV. This algorithm was modified to save the positions and dimensions of the detection rectangles. This information was input into MATLAB where a smoothed image of the number of rectangles was created and was thresholded in the detection process. For the results shown here, each of the cascades was run with all the images and the results in terms of probability of detection and false alarm rate are shown as a function of the 2 seabed features. The mean feature values for each sonar image were used to associate the image with a box in two-dimensional feature space and the average probability of detection and false alarms/file are computed within the box. For the results shown here, the dimensions of the boxes grow logarithmically. In Fig. 3a the false alarm rate (per original sonar tile size) is shown for Cascade 2 for a fixed rectangle threshold of 8. In Fig. 3b the corresponding probability of detection is shown. This varying probability of detection is computed by considering the total number of ground truth detections which fall within a feature-space tile and the actual number detected. The number of detections made within a window may be statistically small

and there are regions of the feature space which are not sampled by the detections. For the results shown we consider only tiles for which there should be at least 3 detections (for the probability of detection displays) and tiles in which there are at least 10 files for the false alarm results. In Figs. 3c and 3d, the results for Cascade 1 are shown. Cascade 2 was trained with the clutter images corresponding to the seabed features of the green area of feature space in Fig. 2a, whereas Cascade 1 is trained against the clutter corresponding to the red section. It can be seen for the region of low feature values (benign seabeds), the performances of the 2 cascades are almost identical. However, for Feature 2 greater than approximately 23, the performance of Cascade 2 is superior in terms of both detection and false alarm rates. For the region, - described by a feature 2 value of approximately 15-20, and feature 1 - less than 2, Cascade 1 is superior in terms of the false alarm rate which is consistent with its training set. Both of these cascades, which were trained to discriminate against complex backgrounds, performed well with the benign seabeds.

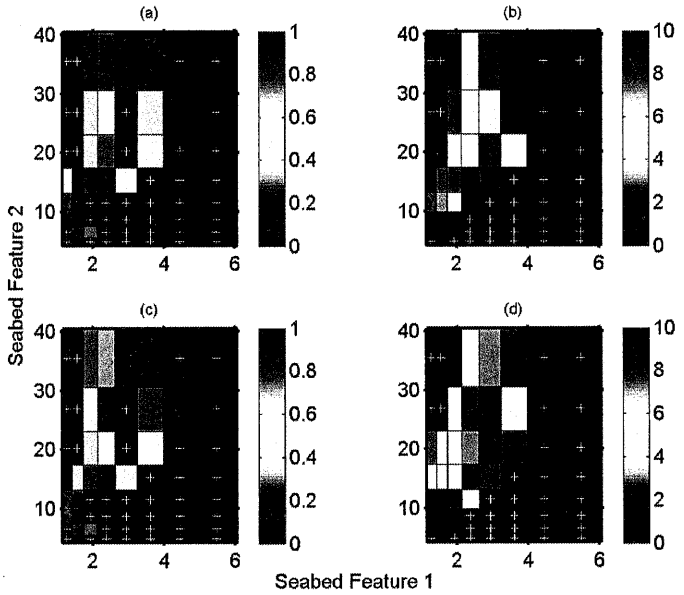

Figure 3 (a) averaged detection rate for Cascade 2 (b) averaged false alarm rate/file for Cascade 2 (c),(d) the same as (a),(b) but for Cascade 1. Results shown for a rectangle threshold of 8. The rectangles with + symbol indicate that there is not sufficient data.

As discussed in Section 3, it is possible to describe the performance/feature dependence via a model, in this case, an ε-SVR model. The construction of the regression model used an unbinned version of the CMRE COLOSSUS II data set. This work resulted in a three feature tuple, where each tuple included the two primary environmental features, and the related threshold for detection. Each tuple was associated with the appropriate label, the false alarm rate, to build the supervised training set. The training set consisted of relatively small number of feature vectors and labels (5690 out of a possible 99280). However, the training set was chosen such that there was a fairly equitable number of samples from the densely and sparsely populated areas of feature space. Although various kernels were evaluated including linear, polynomial and radial basis function (RBF), the RBF kernel was found to provide the highest level of accuracy.

Although SVR models have been shown to be capable with both non-sparse and sparse training data, careful preparation of the data- and selection of the parameters for the SVR are required. These parameters include the ε parameter for determining the training samples which are used to tune the model, the cost parameter, which trades off the smoothness of the function with the potential for misclassification, and γ, a parameter describing the width of the RBF kernel used. These parameters are typically implementation dependent, and require specific tuning and training. In the current work, the data set was separated into training and testing sets, with each being normalized before use. To ensure that valid parameters have been used, a grid search method over

the parameter space (ε,C,γ) was conducted to fit the most likely parameters, using five-fold cross validation for fitness. The parameters selected for this work (ε,C,γ) were (0.28, 2.4, 44). In Fig.4a, we show the data (false alarm rate/file for a rectangle threshold of 8) binned at a much finer resolution than that of Fig.3b and with no constraint on the number of images per cell. The logarithm of the feature values are used and the samples were scaled to lie in the interval [0,1]. In Fig.4b, the predicted values of the false alarm rate f are shown for a regularly gridded set of "test" feature values. The resulting predicted smooth surface is consistent with the results of Fig.4a. It should also be noted that the model is learned for other values of the threshold feature. There are regions in Fig.4b where there is little or no data in the training set and these regions should be used with caution.

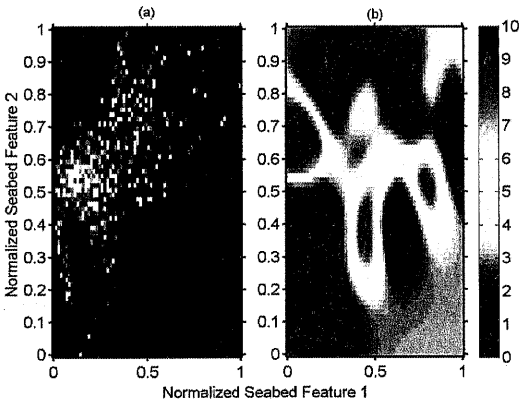

Figure 4 (a) high resolution binned averaged false alarm rate for Cascade 2 (b) trained regression model of averaged false alarm rate.

We now apply the cascades, trained with the COLOSSUS data, to the MANEX 13 data set. The MANEX 13 data is processed in the same manner as above. The resulting performance for the cascades 2 and 1 are shown in Fig.5. The distribution for the 2 seabed features for the MANEX site are for the most part included within the COLOSSUS Site distributions. This does not mean that there are not new types of seabed, only that the associated seabed features are within the span of those from the COLOSSUS site data. The results of Fig. 5 correspond approximately to those of the lower quadrant of Fig. 3. They are reasonably consistent with those from the COLOSSUS Site (Fig.3), a high detection rate (particularly for Cascade 2) and a relatively low false alarm rate for Feature 2 values less than approximately 15. In this region of feature space, the false alarm performance is better than for the COLOSSUS site. The higher false alarm rates for f2 approximately equal to 20 (particularly Cascade 2) are likely due to images with Posidonia.

We can use the empirical performance data (or the learned model) to predict other ATR parameters of interest. The required ATR threshold to obtain a specified false alarm rate can be estimated in the following manner. Within each cell, the number of false detections is considered as a function of the rectangle threshold. Two thresholds will bracket the desired false alarm rate. Then inverse linear interpolation can be used to determine a value (generally, a non-integer number) corresponding more precisely to the false alarm rate. In Fig.6a and 6c, the required thresholds are shown for Cascade 2 for the COLOSSUS and MANEX sites. As would be expected, for both sites the threshold is quite low for benign seabeds. For larger values of seabed feature 2, the threshold increases considerably, particularly for the COLOSSUS site. A lower threshold is required for the MANEX site at these values. Thus, in this region of feature space, the parameter prediction from the COLOSSUS site was found to be much too high. We believe that the reason for this is that some of the COLOSSUS files in this region of feature space contain minelike clutter, i.e., rocks which yield a high rectangle count. Histograms of the rectangle count distribution show that, indeed, the COLOSSUS site data has more incidences of files with multiple high rectangle detections. In Fig.6b, the required threshold is shown if we reject files which contain detection with 70 or more rectangles. Then, as can be seen, the predicted threshold values appear to be much closer to those of the MANEX site. This is an indication that using just 2 seabed features is not always sufficient to characterize performance.

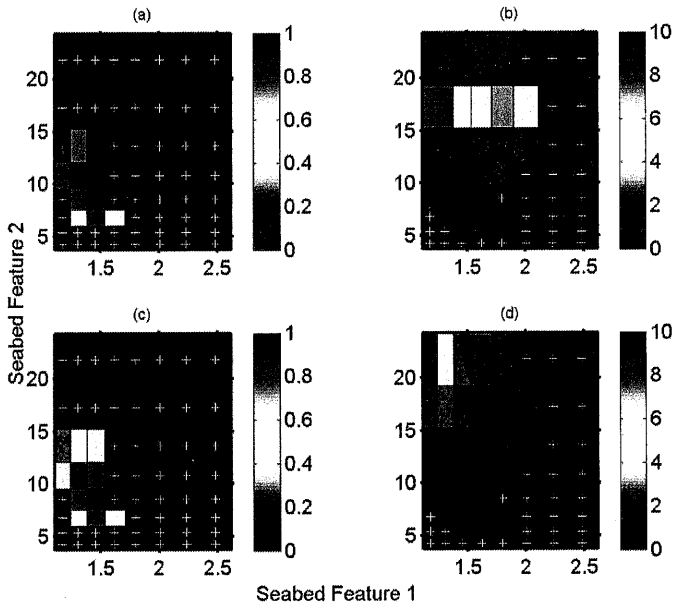

Figure 5 (a) averaged detection rate for Cascade 2 (b) averaged false alarm rate for Cascade 2 (c),(d) the same as (a),(b) but for Cascade 1.

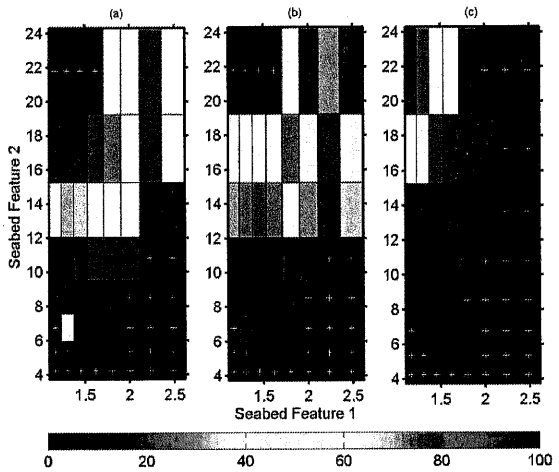

Figure 6 (a) the threshold required for COLOSSUS set for a false alarm rate of 1/file (b) also for the COLOSSUS set, but only considering files not containing detections with more than 70 rectangles (c) required thresholds for MANEX site.

6 CONCLUSIONS AND FUTURE WORK

In this work we have shown an initial method for using sensed environmental features to predict ATR performance. Although we have applied this method to a cascaded detector, the work can be extended to other detector/classifier methods such as a matched filter. As with all trained methods, generalization is the critical measure of performance. The utility of the trained method is in its ability to effectively detect targets in a previously unseen area. Using the aforementioned data sets, we have shown the effective generalization performance of the trained cascade detector. The predicted performance as a function of seabed features also generalized relatively well, however there were regions in feature space where the performance predicted by the COLOSSUS II data set was significantly poorer than the MANEX data set. This was observed specifically in the higher complexity regions for the required threshold level to maintain a fixed false alarm rate. Here, we

argue that another feature, one which describes the complexity of the seafloor from a mine like object perspective is required. Even with the Colossus II data, there was a high variance in performance for very close feature values. This suggests that another feature value will assist in effectively making the performance/feature function more separable, and further increasing generalization performance of the prediction models.

This initial work shows promise in the prediction of ATR performance for an unseen area using simple environmental features. This provides a technique for adaptively tuning an ATR system allowing an *in-situ* method for increasing ATR performance. It would also provide an *in-situ* estimation of probability of detection performance for a given threshold. An ideal example of where adaptive processing is useful is in the AUV case, where human interaction is not possible. Future work for this technique includes an on-line monitoring method for false alarm rate, selection of a new environmental feature, and the collection and use of a larger training set. As noted above, the AUV case is a highly useful application of this method, to ensure effective performance during the mission. This monitoring system would not only collect features for tuning the ATR, but also monitoring actual vs predicted performance for unseen areas. To improve the prediction performance, a new feature will be incorporated into the model to more effectively describe the environment from an MCM perspective. Finally, as noted in Section 4, the training set was very biased in that it was dense for specific feature values, and highly sparse for others. In particular, the number of images with targets was relatively small, resulting in large areas of the feature space for which there was no data. The collection of a broader training set which contains more samples for the sparse areas will improve the accuracy of the prediction models.

7 REFERENCES

1. D. Williams, "On adaptive underwater object detection," *in Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)}* San Francisco, U.S.A, 4741--4748, (Sept. 2011).
2. D. Williams and E. Fakiris, "Exploiting Environmental Information for Improved Underwater Target Classification in Sonar Imagery", IEEE Trans. on Geoscience and Remote Sensing, to be published, 2014.
3. A. Lyons, D. Abraham, J. Groen, and W.L.J. Fox, "Statistics of template-filtered synthetic aperture sonar images", *in proceedings of UAM 2011,* Kos, Greece, 2011.
4. J. Gazagnaire, J.T. Cobb, and P.P. Beaujen, "Environmentally-Adaptive Automated Target Recognition Algorithm Using Parametric Characterization of Seabed Textures", *in Proceeding of IEEE/MTS Oceans 2010,* Seattle, Washington.
5. P.Viola and M.Jones, "Rapid object detection using a boosted cascade of simple features," IEEE CVPR, 2001.
6. Y.Petillot, Y.Pailhas, J.Sawas, N.Valeyrie, and J.Bell, "Target recognition in synthetic aperture sonar and high resolution side scan sonar using auvs," in Proceedings of International Conference: Synthetic Aperture Sonar and Synthetic Aperture Radar, Institute of Acoustics Proceedings, Lerici, Italy, (Sept. 2010).
7. O. Daniell, Y. Petillot, and S. Reed, "Unsupervised Sea-Floor Classification for Automatic Target Recognition*", in International Conference on Underwater Remote Sensing, 2012.*
8. E. Fakiris, D. Williams, M. Couillard, and W.L.J. Fox, "Sea-Floor Acoustic Anisotropy and Complexity Assessment Towards Prediction of ATR Performance" *in Proceedings of First International Conference and Exhibition on Underwater Acoustics,* Corfu, Greece, 2013.
9. T. Sams, J. Hansen, E. Thisen and B. Stage, "Segmentation of sidescan sonar images", Technical Report of the Danish Defence Research Establishment M-21, 2004.
10. V. Vapnik, S. Golowich and A. Smola, "Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing", Neural Information Processing Systems, Vol. 9. (1997)
11. C. Chang and C. Lin, "LIBSVM: A library for support vector machines", ACM Transactions on Intelligent Systems and Technology, Vol. 2(3), 1-27. (2011).
12. Open Source Computer Vision OpenCV 2.4.3 , http://opencv.willowgarage.com/wiki/, (Access date:January 2013).