Computer-aided detection and classification of minelike objects using template-based features

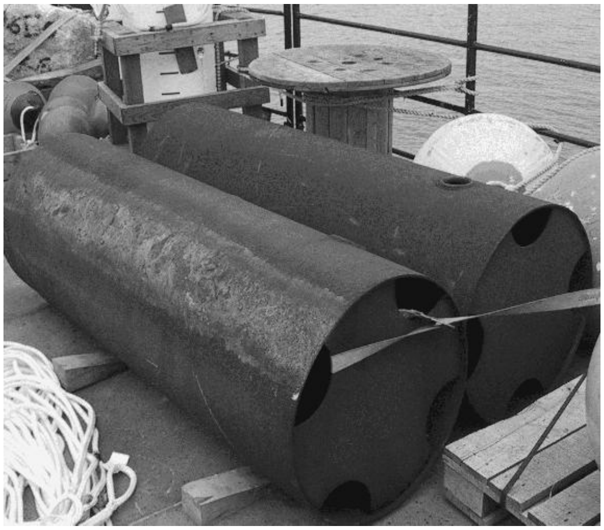
John A. Fawcett,
DRDC-Atlantic,
P.O. Box 1012, Dartmouth, N.S.,
Canada, B2Y 3Z7

***Abstract* - In this paper the classification of small sidescan sonar images is considered. The features of the image that are used for classification are the outputs from various convolution masks based on ray-generated images of targets and their corresponding edge masks at various aspects and ranges. This procedure has the advantage that a detailed segmentation of the image into highlight and shadow is avoided. The procedure is efficient to implement and robust. In this paper we concentrate on the identification of cylindrical objects. This case is more complicated than some other targets because of the aspect dependence. A set of sidescan sonar images of a specific type of cylinder, as well as images from a large variety of clutter events is considered. A subset of the cylinder images is used for training the algorithm.**

I. INTRODUCTION

The classification of objects in sidescan sonar imagery usually consists of two basic stages. The first is feature extraction. This usually involves segmenting the small image of interest into shadow, highlight, and background regions. Then various features related to the size, shape, orientation, and statistics of the highlight and shadow regions are defined. These features are then used by a classification procedure to classify the object as clutter or minelike. A further breakdown of these classes may also be performed. A problem with such an approach often lies with the accuracy of the feature extraction. For example, the segmentation of an image into well defined highlight and shadow regions may involve sophisticated algorithms [1-2]. Large amounts of speckle and highlight features within the shadow from, for example, internal ringing of the target, may cause problems for some segmentation algorithms. Another problem that arises with standard sidescan sonar imagery is that there are a variety of practical reasons for which an image of a target will not have correct physical dimensions. For example, the target may be partially buried or tipped over slightly which will skew the shadow lengths. If ungeoreferenced imagery is used, pitch and yaw variation can cause distortions of the image. Thus, it is important to have some robustness in the classifier. Another important consideration with sidescan sonar imagery is that certain clutter events (e.g., rocks of the appropriate dimension) will look very minelike and it may be better to have them classified as such rather than attempt to train a classifier to reject them.

The approach of this paper is to define features based on the output of template convolutions. This is similar to the approach of [3] which used the output of a template convolution as a detection method. In this paper, templates for a cylinder of interest are precomputed using raytracing for different aspects and ranges. The maximum correlation of these templates with a normalized version of the sonar image is one feature. But in addition, we also consider the convolution of edge masks, different normalizations of the image, enlarged versions of the target template, etc. to define a set of 10 features. We then compute the lower bounds on these outputs for a set of known targets and then subsequently use a threshold based upon these bounds to determine whether an image is a cylinder or not. It is anticipated that there will be certainly be false alarms using this approach. However, it is hoped that the images for these cases will visually resemble a cylinder at some aspect. This work is also related to that of [4] where the concept of deformable templates is used. In this method a target is hypothesized and then its template is allowed to deform in an iterative fashion to produce a segmentation of the image. A measure of the fit of the template to the data can be used to classify the target. Another paper related to the present work is [5] where classification was performed using the images' pixel values. The number of features was first reduced by using both principal component and linear discriminant analysis. The discriminating features between object classes could be interpreted as the outputs from the convolutions of discriminating templates. From this paper, it is clear that the outputs from convolving the image with templates can yield valuable classification information. However, [5] used only simulated data and restricted the images to one cross range.



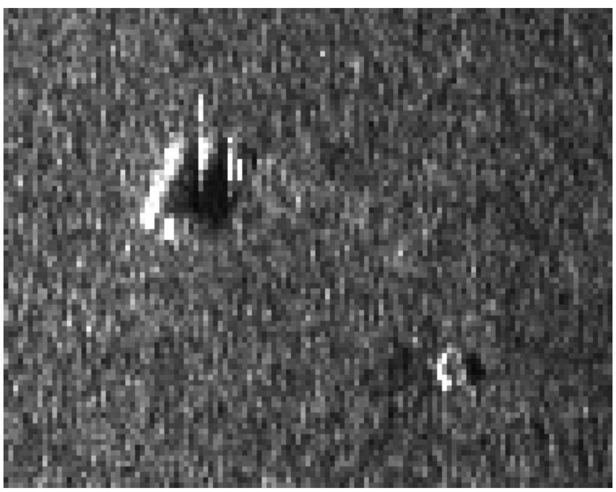


Fig. 1. Two of the cylinders used in trial and a sidescan sonar image of one of the cylinders - note the ringing features

II. DATA SETS

During June 2001, 4 cylinders were deployed in the area of Herring Cove in Halifax Harbour. These cylinders are 1.8 m long and 0.6m in diameter. A photograph of two of these cylinders and a corresponding sidescan sonar image is shown in Fig.1.

These cylinders were free-flooding and evidence of internal scattering of the sonar energy within the cylinder was often strongly evident. This data and all the data for this paper were collected with a Klein 5500 sidescan sonar using its 10-cm along-track resolution. The Herring Cove area has a predominantly sandy bottom but there are also regions with rocks, ripples, etc. which provide clutter events for this study. A series of north/south, east/west and some diagonal runs were performed with the Klein 5500 sidescan sonar. The set of cylinder images obtained with this initial survey will be used to compute various statistics for the template matching. Another series of runs along diagonal tracks were performed for the same targets. The survey was carried out from a small boat and, at times, there are various pitch and speed artifacts observed in the sonar imagery. In addition this site was returned to during the DRDC-Atlantic/SACLANTCEN MAPLE 2001 trial [6] at which time the sonar was deployed from a much larger ship, the CFAV Quest. The images (cylinders and clutter) from the second set of diagonal runs as well as those from the MAPLE 2001 (Herring Cove) trial will form the “testing set”. In addition we chose many clutter events during the time of the “training set” and we will see how well these are rejected. It should be emphasized that we do not train with any clutter events. We simply compute bounds for the feature values of the known targets and use thresholds based upon these bounds to classify the other images. The Klein files were read into a program, which performed a very rough automated detection by performing a simple convolution with a template, similar to that described in [3]. The detection threshold for the output of this filter was set to a low level so that many detections were made. These detections were then manually inspected and all the cylinder images and selected clutter events were output. The output files were the Klein beam data for 21 pings centred on the object as well as the associated navigational/orientation data. For the tow speeds of this trial, approximately 4 knots, the Klein 5500 produce a significant number of redundant beams. That is, certain sets of the beams are deemed as “visible”- required to make a contiguous image and “redundant” – those which overlap the coverage provided by the visible beams. The towfish ping rate is typically adjusted according to the estimated tow speed so that the set of visible beams will form, as nearly as possible, a contiguous, non-overlapping set. In this paper we just use the visible beams to form an image. However, in the future, it would be interesting to include the redundant beams and to account for towfish motion (i.e. speed and pitch variation) to improve the image quality [7] and determine any difference in subsequent classification performance.

III. Ray Generation of Templates

We generate a template of what the highlight, shadow, and background regions for a given target should look like by using a simple object/ray-tracing code. The cylinder is specified by a set of triangular facets. The cylinder model can easily be rotated by applying a rotation matrix to the facets' nodes. The concept of using ray-tracing to interpret sidescan images and as a possible classification tool is not new and can be found in, for example, [8] and [9]. For the cylinder we compute the templates for 13 azimuthal angles (-90 degrees to 90 degrees), 16 ranges (15 to 80m) and 2 different-sized cylinders (0.3 m radius, 1.8m long) and (0.3 m radius 2.5m long), the first one being the one which corresponds to the cylinder we are looking for. The template computations are done for sonar altitudes of 12, 15 and 18m. The along track spatial and across-track temporal spacings correspond to those of the Klein 5500 sidescan sonar (high resolution mode). It is certainly possible that one could use a subroutine call to the ray-generation code to compute the exact template for a specific altitude and cross-range. One could also try a more exact azimuthal match than we do in this paper. However, for this paper, we took the approach of precomputing the templates and then accessing them as needed. During the actual deployment, the altitude of the sonar is recorded and will not be equal to one of the computed altitudes. We use the set of templates from the nearest altitude and the range, which gives approximately the same grazing angle as the true altitude. In Fig.2 we show 4 representative templates at 4 different aspect angles for a cross-range of 55m and altitude of 15m. The amplitudes of the highlight are somewhat adhoc – we simply used the cosine of the angle between the ray and the surface normal to compute an amplitude for a ray. A template is normalized so that its largest amplitude is unity. The shadow values are set to negative one. One could certainly include more realistic factors to try to account for diffractions from edges, the beamwidth of the sonar, etc.

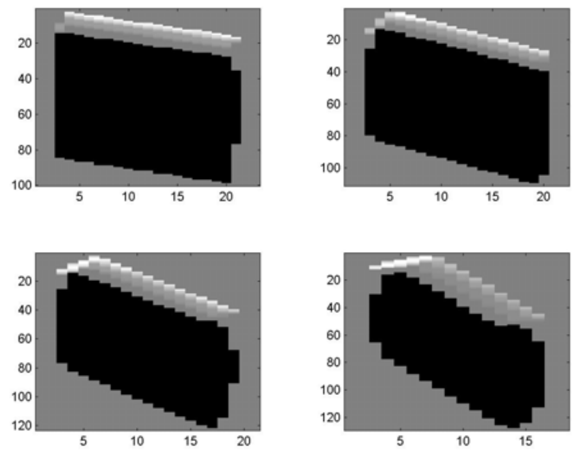


Fig. 2. Sample templates of the cylinder at different aspect for fixed altitude (15m) and cross-range (55m)

IV. FEATURE GENERATION

The sets of data which were manually extracted from the Klein images yield a set of "swath" files and associated information including the time index for the manual detections. A small box of data about each detection is then extracted; in particular, 231 time points X 21 pings are extracted. This is then reduced to just the visible beams. The overall backscatter levels of the sonar depend upon the range of the target and also the characteristics of the seabed. In order that various feature values should not depend upon these variable conditions, it is necessary to carefully normalize the data. The data values of the small image are first clipped at a factor of three times the median value and then scaled between 0 and 255. The mean of this data is then subtracted from the image so that highlight regions are positive and shadow regions are negative. In order to denote highlight and shadow values, we now use simple thresholding. In general, we sort the data values and define anything above a certain threshold (e.g., the top 90%) as "highlight" and below a certain threshold (e.g., lower 20%) as "shadow". This is similar to the normalization applied in [10]. This simple thresholding when applied to the small data sets may sometimes have problems. For example, in areas of poor background/shadow contrast the resulting shadow segmentation may not be accurate. However, in general this approach worked well. Also, since the method of this paper only requires some approximate measure of the shadow and highlight regions, it is not crucial to have highly accurate segmentations. For this data set we also applied a 9-point median filter in the across-track direction in an attempt to minimize some of the "ringing" effects, which were present in many of the cylinder images. In fact, one could use these "ringing" features as a very strong classification clue. However, since this a result of using free-flooding cylinders, we did not wish to do this and consider the ringing to be a "nuisance" in the images. It is a factor which can cause problems with image segmentation algorithms and it also causes shadow and highlight mismatch in the template matching of this paper.

The next step in the feature extraction procedure is to compute the matching between the different templates and the given image. Since we know the range of the detection, it is possible to select the appropriate templates for that range and altitude. We then consider the different azimuthal orientations and determine the optimal match in terms of the correlations. For this correlation we normalize the template by its L_2 norm. Simply determining the largest template correlation with the image usually determines the appropriate aspect. However, in a few cases where there is poor shadow structure and perhaps significant ringing, we found that the variation of the correlation with aspect could have more than one maximum. In these cases there were two local maxima of almost equal value. Another form of correlation is to first take the absolute value of the normalized image and the template. In this way the target highlight and shadow both have non-zero values and similarly for the templates We found that by optimizing the sum of the template/data correlation and some fraction of correlation of the absolute values helped stabilize this process (we used a factor of 0.2). There was one cylinder image in the training set with significant ringing and a small amount of shadow for which the determined azimuth did not agree well visually with what one might select. However, it was definitely the maximizing template and we found that the resultant computed features were consistent with those from the other images so we kept it in the training set.

The values of the templates/image correlation are two important features, but alone are not enough to distinguish the cylinders from all the clutter. For example, one could consider a very large clutter image, which would produce a high correlation but because it is, in fact, much larger than the cylinder it would be obvious, visually, that this image is not a cylinder. In order to constrain the acceptable size of objects, we consider the difference operator in the column and row directions, for both the image and the template, to compute edge masks. The correlations between the edge masks provide 2 more features. Further, we also consider a cylinder which is 2.5m long and consider its correlation with the image. The logic here is that any relative increase in correlation over using the template with correct length should not be significant. This feature is the ratio of the correlation using the 1.8 and 2.5 m templates. In Fig. 3 the edge masks for the image and the corresponding edge templates. Besides these features we compute a number of other template features. For example, the percentage of a templates shadow points which correspond to those of the image, the correlation value of highlight regions, the correlation of the modulus of the gradient, etc. We use a total of 10 features. There are 28 cylinders in the first 76 images which are used to compute the statistics of the template features. The cylinders correspond to a variety of aspects, cross-ranges and towfish altitudes. In computing the features, we are careful to normalize by the size of the template, so that the values of the features should be approximately independent of the estimated aspect and cross-range.

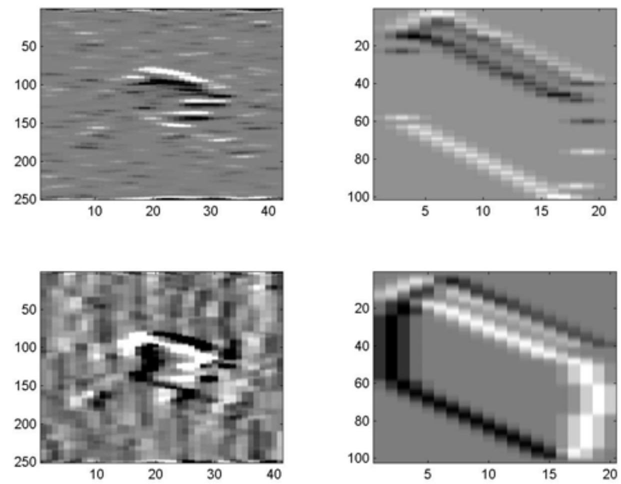


Fig. 3. Difference masks for one of the cylinder images (column 1) and template (column 2) in the across-track and along-track directions.

V. NUMERICAL RESULTS

We first consider the 28 cylinders from the training set of Herring Cove. We then compute the various features for these cylinders and compute their lower bounds. The data set containing the 28 cylinder images also contained 48 clutter events. We will denote this combined data set as Data Set 1. In Fig.4 the first 16 cylinders from this set and the estimated azimuthal templates are shown. It is important to note that when using the training cylinders, the azimuthal matching is still numerically done, so that the computed feature values reflect any inaccuracy in this portion of the procedure. In Fig.5, we show, as an example, the histograms for Features 1 and Feature 5 for just the cylinders of Data Set 1 and for all the images of Data Set 1. Feature 1 is a measure of the correlation between the optimal azimuthal template and the data (normalized by the L_1 norm of the template) and Feature 5 is a measure of the correlation between the along-track difference of the template and the along-track difference of the data (normalized by the L_1 norm of the template).

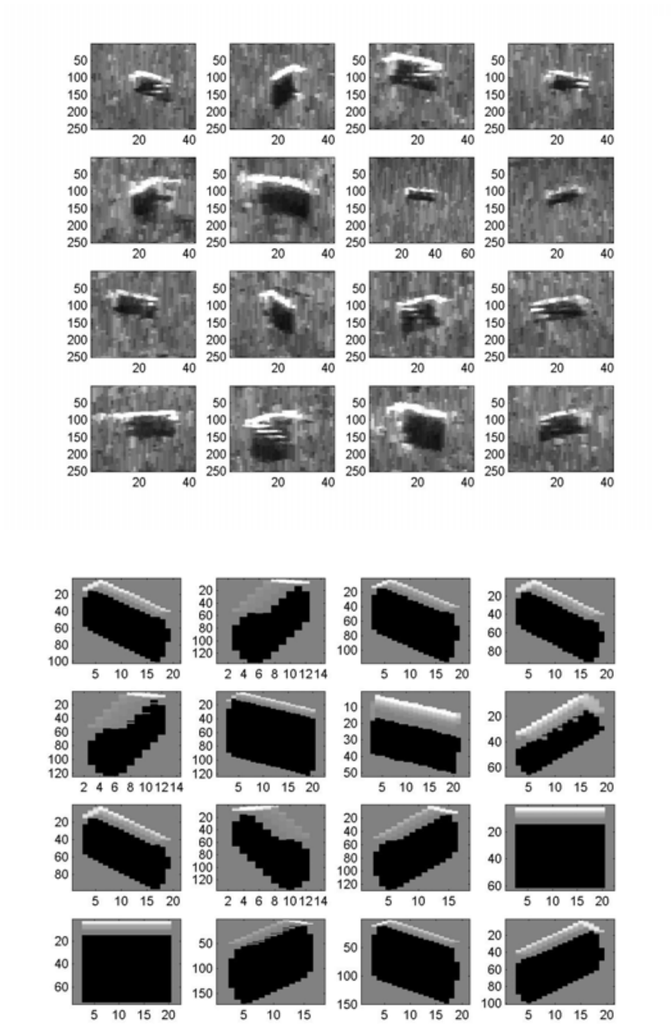


Fig. 4. The first 16 cylinder images used for training and their associated azimuthal templates – for the templates, white represents a value of unity and black (shadow) has the value of negative one.

As can be seen from Fig.5, by setting a lower threshold on these 2 feature values, one can eliminate several of the clutter events. By using this procedure for all the features, a significant fraction of the clutter images can be rejected. There are some clutter events, which are below several of the thresholds; however, it is hoped that by adding additional features an increasing number of clutter events will be rejected. The thresholds are taken to be some fraction of the lower bounds (upper bound for one feature) determined from the training set. These percentages are parameters which can be varied. The lower we allow the thresholds for the features,

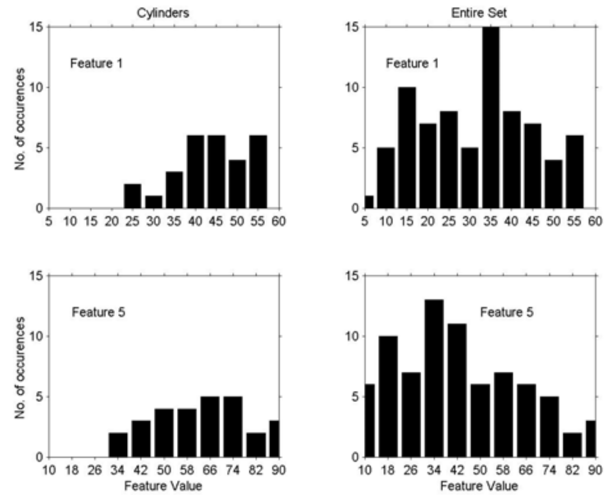


Fig. 5. Histograms for Data Set 1 for (a) just the cylinders (b) the entire data set

the higher the probability of detection but also the higher the false alarm rate. For all the features, except for Feature 5 as discussed below, we found that 80% of the lower bounds and (120% of an upper bound for one of the features), computed from the training set worked very well with the testing set. The data set, Data Set 2, was constructed from using the data from the latter part of the day of June, 2001 as well as the data from the MAPLE trial at the Herring Cove site. This set contains 56 images, 29 of which were classed as cylinders and 27 as clutter. Several of the cylinders in this set were endon and we found that one of them failed the bound on Feature 5, which is related to the difference operator in the along-track direction. By decreasing this bound to 70% of the lower bound as computed from the training set we correctly classified this cylinder with no increase in the amount of false alarms, so this is the bound we use for this particular feature for the results of this paper. It is reasonable that for endon cylinders, which have a narrow highlight and shadow area, this feature may be problematic. For the 29 cylinders in Data Set 2 the method of this paper found 27 and missed two. In Fig. 6 we show the first 16 cylinders detected. As can be seen in Fig. 6 a wide variety of different cylinder aspects at different cross-ranges and for the towfish at different altitudes have been successfully classified. The endon targets (image 2,5,6,7) in Fig.6 have a complicated highlight structure with ringing effects in the shadow. It is encouraging that the method is sufficiently robust to correctly classify the cylinders under these conditions. For the 2 undetected cylinders, it is understandable why these 2 particular images were not correctly classified. The first one was thought to be an endon cylinder, but is at short cross range and the recorded altitude of the towfish was 22 m. The cylinder of the second misclassification appears significantly shorter than it should be. This may be a result of the fact that the "visible" beam set did not cover the cylinder appropriately (and for this case the estimated ship speed was less than 2 knots). It is, of course, possible to lower the feature thresholds, so that even these 2 cylinders are classified; however, in this case the number of false alarms (clutter images) increases. Of the 75 clutter events from Data Set 1 and Data Set 2, 13 were classed as cylinders, - meaning that 83% of the clutter images were rejected. In Fig.7 we show the 13 clutter images incorrectly classed as cylinders. Many of these images do appear to be cylinder-like. The smaller images of images 10,12, and 13 correspond nicely to near endon cylinders at short range. Some of the images of Fig. 7 might be rejected by a trained human operator. However, the template match is reasonable, in terms of matching up much of the shadow, highlight, and edge structure of the image. The method seems to provide a good clutter rejection rate while detecting a high percentage (27/29 = 93%) of the cylinders. The rejection rate could, perhaps, be improved by including additional features. In particular, we found that the detailed structure of the highlight region of a rock might appear "visually" different from that of the cylinder, even though the general size, orientation might fit within the bounds imposed by the feature set. Deriving features to capture this behaviour is an area of further research. However, the highlight region is difficult to model well, as its behaviour can be somewhat unpredictable.

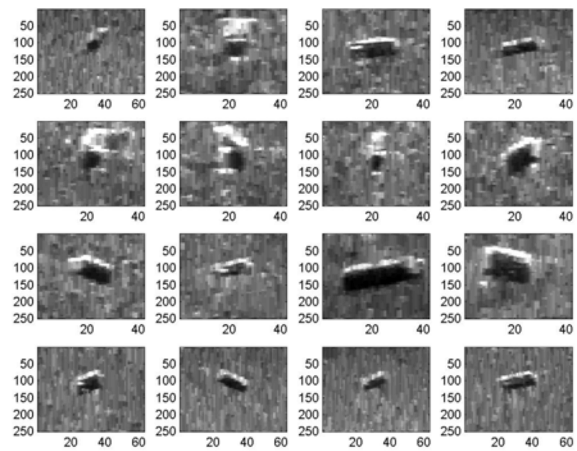


Fig. 6. The first 16 detected cylinders - in the text they are considered numbered from left to right, starting at the upper left corner

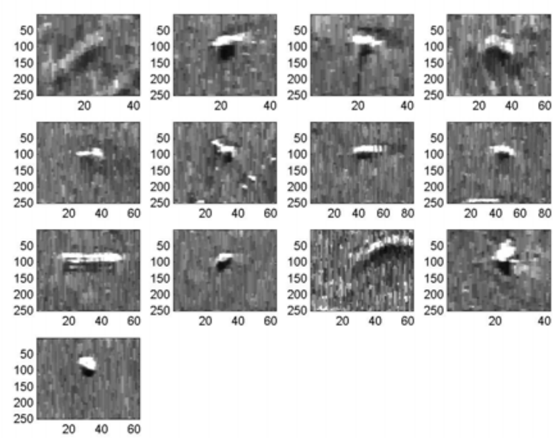


Fig. 7. The 13 clutter events misclassified as cylinders

VI. SUMMARY AND DISCUSSION OF RESULTS

We have presented a method for classifying objects using template-based features. These features are based upon the correlation of a ray-based model of highlight shadow structure, appropriate for the detection range and altitude, with a normalized version of the sonar image. First, the optimum azimuth of the template is determined and then a variety of templates, including discrete derivatives of the template and the image, are used to compute a set of features. Using a set of known cylinders, lower bounds for these feature values are determined (for one feature an upper bound is used). A threshold for each of these features is then used to decide whether an image is a cylinder or not. The overall detection rate for the cylinders in the testing set was high – 93% and these included a wide variety of cross ranges, altitudes, and aspects. The rejection rate of the clutter events was about 83% and most of the clutter events falsely identified as cylinders were reasonable. Some of the cylinder images had relatively poor highlight and shadow regions because of the ringing effects (due to the water fill) and distortions due to towbody motion. This resulted in many of the computed features having fairly low bounds. Thus this data set provides a good test of the robustness of this approach.

We have not exhaustively tested all possible features or combinations of features in this paper and as mentioned, some additional features regarding the highlight region might improve the false alarm rate. Also, although we have emphasized template-based features, in practice, one could combine this classification method with features obtained from other techniques. As a benchmark case, we have considered only one specific class of targets in this paper. For more than one class, one could construct a set of templates for each class and repeat the approach of this paper. However, the optimal set of features for discriminating a particular class from clutter might vary.

REFERENCES

[1] S. Reed, Y. Petillot, and J. Bell, “Unsupervised mine detection and analysis in side-scan sonar: A comparison of Markov Random Fields and statistical snakes” *in Proceedings of CAD/CAC 2001, Halifax, Nova Scotia, 2001.*

[2] V.L. Myers, “Image segmentation using iteration and fuzzy logic”, *in Proceedings of CAD/CAC 2001, Halifax, Nova Scotia, 2001.*

[3] G.J. Dobeck, J.C. Hyland and L. Smedley, “Automated detection/classification of seamines in sonar imagery” *Proc. SPIE-Int. Soc. Optics,* vol. 3079, pp.90-110, 1997.

[4] I. Quidu, Ph. Malkasse, G. Burel, and P. Vilbe, “Mine classification based on raw sonar data: An approach combining Fourier descriptors, statistical models and genetic algorithms” *OCEANS MTS/IEEE Conf. And Exhibition,* Vol. 1, pp.285-290, 2000.

[5] J.A. Fawcett, “Image-based classification of side-scan sonar detections” *in Proceedings of CAD/CAC 2001, Halifax, Nova Scotia, 2001.*

[6] M.A. Trevorrow, A. Crawford, J. Fawcett, R. Kessel, T. Miller, V. Myers and M. Rowsome, “Synopsis of survey data collected during Q-260 MAPLE 2001 sea-trials with CFAV Quest and NRV Alliance”, DRDC-Atlantic TM 2001-172, November, 2001.

[7] J.A. Fawcett “Processing and displaying multibeam sidescan sonar data: visible, redundant and overlapping beams”, DRDC-Atlantic Technical Memorandum TM 2002-003.

[8] J. Bell, *A model for the simulation of sidescan sonar,* Ph.D. Thesis, Herriot-Watt University, August 1995.

[9] B. Zerr, “Automated mine classification approach based on multiple aspects acquired by an autonomous underwater vehicle” *presented at CAD/CAC 2001, Halifax, N.S., 2001.*

[10] S.G. Johnson and M.A. Deaett, “The application of automated recognition techniques to sidescan sonar imagery” IEEE J. Oceanic Eng., Vol.19, pp.138-144, 1994.