False-alarm reduction in mine classification using multiple looks from a synthetic aperture sonar

J. Groen, E. Coiras & D. P. Williams
NATO Undersea Research Centre
Viale S. Bartolomeo 400
19126, La Spezia, Italy

***Abstract*- An algorithm is proposed for performing underwater mine classification when provided with multiple synthetic aperture sonar (SAS) views of each object. The proposed method consists of three main steps. First, the multiple images of a given object are registered to a common reference frame. Second, the registered images are fused to form a single image, thereby allowing all of the information contained in the multiple images to be exploited simultaneously. Third, the fused image is compared to a library of image-templates of targets of interest. The approach is demonstrated on one data set of simulated SAS images and on one data set of real, measured SAS images collected at sea. As the number of views of an object increases, classification performance improves and the dependence on the orientation of the object lessens.**

I. INTRODUCTION

The detection and classification of underwater mines is an important task with strong implications for the safety and security of ports, harbors, and the open seas. In recent years, the ability to successfully address this challenge has improved with the significant advances made in synthetic aperture sonar (SAS) imaging systems. An autonomous underwater vehicle (AUV) equipped with a state-of-the-art SAS system can provide imagery of the underwater environment with resolution on the order of a centimeter while achieving an area coverage rate of over a square kilometer per hour.

Nevertheless, even with the advent of the higher resolution SAS systems, the task of mine classification when provided with only a single view (or "look") of an object remains a challenging enterprise. This difficulty can be attributed to the relative similarity in appearance between mines and benign clutter objects in SAS imagery. For example, if a mine is observed at a certain orientation, it may be difficult to distinguish it from a rock. However, viewing the mine from a second orientation may reveal previously obscured characteristics that differentiate it from a rock. Fig. 1 provides an example of this case.

In general, the information accrued from multiple views of an object should translate into a reduced false alarm rate, and hence, improved mine classification performance. This basic concept motivates the work contained herein regarding the collection of multi-view data for the classification of underwater mines.

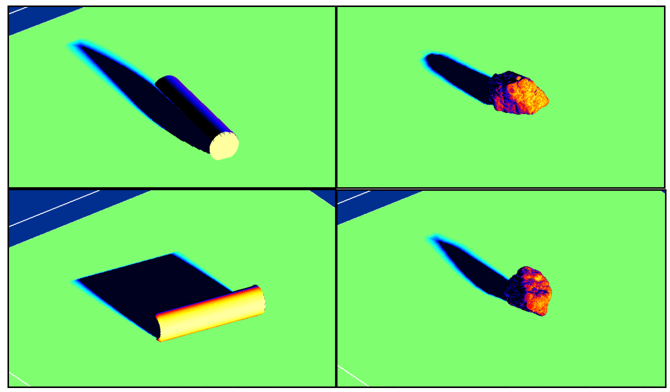

Figure 1. Potential classification confusion between a cylinder and a rock if only one view is provided. Two simulated SAS images each, at different orientations, of a cylinder (left side) and a rock (right side).

Attempting to perform underwater mine classification using multiple SAS views is a relatively immature area of research. One approach [1] was based on finding the single highest maximum correlation between (i) a set of views (i.e., images) of a training shape of interest and (ii) a set of views of a given testing object. Classification was then performed by using this measure of similarity, termed the *affinity*, directly.

A second approach [2] employed a partially observable Markov decision process (POMDP) to perform multi-view classification of underwater objects.

In this work, we take a different approach that is based on fusing multiple views of an object into a single image. Classification of the object is then performed by comparing the fused image to a library of simulated image-templates of targets of interest.

This approach is particularly well-suited for the underwater mine classification problem for several reasons. For one, it fully exploits the recent advances in SAS systems by focusing on the detailed shape information of the objects that the high-resolution imagery provides. But more importantly, the approach also overcomes the unique challenges presented by the general underwater mine classification problem that cause standard (single-view) classification approaches to fail. There is also the additional advantage that the data of a contact can be visualised in a natural way for analysis by a human operator.

A fundamental assumption of machine-learning algorithms is that the underlying mechanisms (and hence statistics) that generate the training and testing data are the same [3]. In the underwater mine classification problem, however, this implicit assumption is often violated because different types of clutter

978-1-4244-5222-4/10/$26.00 ©2010 IEEE

objects can be encountered at different sites. As a result, a classifier learned from clutter training data collected at one site often will not generalize well on testing data collected at a different site.

Furthermore, in the underwater mine problem, the universe of mine types that one is interested in detecting and classifying is typically reduced (through common sense and intelligence). However, within this small class of target types, marked differences exist among the different mines. As a result, the features that are salient for certain mines are not relevant for classifying other mines. This fact induces a need for larger feature spaces, which in turn necessitates more training data to build a reliable classifier.

The fact that the image of a mine is highly aspect-dependent further increases the need for more training data (namely, from multiple aspects). For example, the image of a broadside cylinder will look very different from the image of an end-fire cylinder (cf. Fig. 1). In general, the relative paucity of training data — of different targets, at different ranges, at different aspects, and in different site conditions — contributes to the difficulty of building a robust classifier.

The approach proposed in this work combats these challenges inherent to the underwater mine classification task by obviating explicit feature extraction and classifier construction. Moreover, in the proposed multi-view classification method, there is no constraint on the number of views that can be handled. Importantly, the elegantly simple approach is naturally suited to allow each object to be viewed an arbitrary number of times.

Unlike the affinity method in [1], this proposed approach has the advantage that the creation of a single image allows all of the information contained in the multiple images to be exploited simultaneously. However, even with the high resolution imagery provided by the SAS, and the knowledge of the relative target-sensor orientations at which each view is obtained, the manner in which the requisite image fusion is performed is not trivial.

The process employed to perform the image registration, fusion and feature extraction is explained in detail in Section II. Experimental results from applying the proposed multi-view classification method to simulated SAS data are shown in Section III. Experimental results from using real, measured SAS data collected at sea are shown in Section IV. Section V provides a brief analysis of the method's utility for false alarm reduction. Concluding remarks and directions for future work are given in Section VI.

II. Multi-view classification approach

For information fusion of the various sonar views of potential targets on the sea floor such that sea mines are classified more reliably, many approaches can lead to success.

The first choice to be made is to combine the single-view images into one multi-view image or to treat the images separately. For this paper the choice was made to fuse the images into one image, which has two additional advantages for the application to the sea mine classification problem. One advantage is that the product, the multi-view image, which has high-resolution and appears of almost photographic quality and gives an accurate clue to the object's shape, is beneficial for evaluation by a human operator. The other advantage is a by-product of the multi-view image, which is the 3D multi-view image based on shape-from-shading techniques [4], which can be presented to the operator or can be used to extract 3D classification features.

Following the image fusion, the classification processing chain consists of the following steps:

1. Image registration: associating and co-registering the multiple views in order to find a common reference frame.
2. Image fusion: combining the data in the images.
3. Feature extraction: extracting features from the multi-view image designed such that optimal discriminative information for classification is included.
4. Classifier: the actual process that results in the probability that the object is a sea mine based on the feature values obtained.

This article proposes choices for each of the steps. The classifier, based on one classification feature, is kept simple on purpose to avoid training and to better show performance gain for additional views.

A. Image registration

In the processing chain for sonar mine hunting, the classifier is preceded by the detector. Since the system considered here is automated, the philosophy for the detector is that it is a workhorse that does the first sifting and its alarm rate is solely determined by what the classifier can handle in terms of computational load.

The detector may produce a large number of geo-referenced image snippets of possible mines such as shown in Fig. 1a. The objective of the image registration is then to associate all the snippets of the same object even if the geo-referencing is not perfect and the snippets look very different.

Fig. 2a shows four simulated views of a cylinder that was imaged with a synthetic aperture sonar operated at a frequency of 300 kHz with a bandwidth of 60 kHz. The resolution of the images is 20 mm in range r (vertical axis) and 30 mm in cross-range x. The range of the cylinder is 100 m, the altitude z_s of the sonar is 20 m and the cylinder is 2.4 m long and 600 mm wide. The four snippets are 10 m by 10 m, to make sure that both target highlight, target shadow and enough sea floor are contained. In this case the top row snippets are rather similar to the middle row due to the symmetry of the cylinder and the fact that the views are spaced 90 degrees apart.

The single-view images are first converted from slant-range plane to ground-range y using simple geometry:

$$y = \sqrt{r^2 - z_s^2}\,. \qquad (1)$$

For the most part, this conversion does not significantly alter the signature or resolution much (e.g. in this example the target ground range is 98.0 m and the snippet is only stretched by 200 mm after this conversion). Then the images are rotated to enable the overlay based on the orientation of the sonar, which is known.

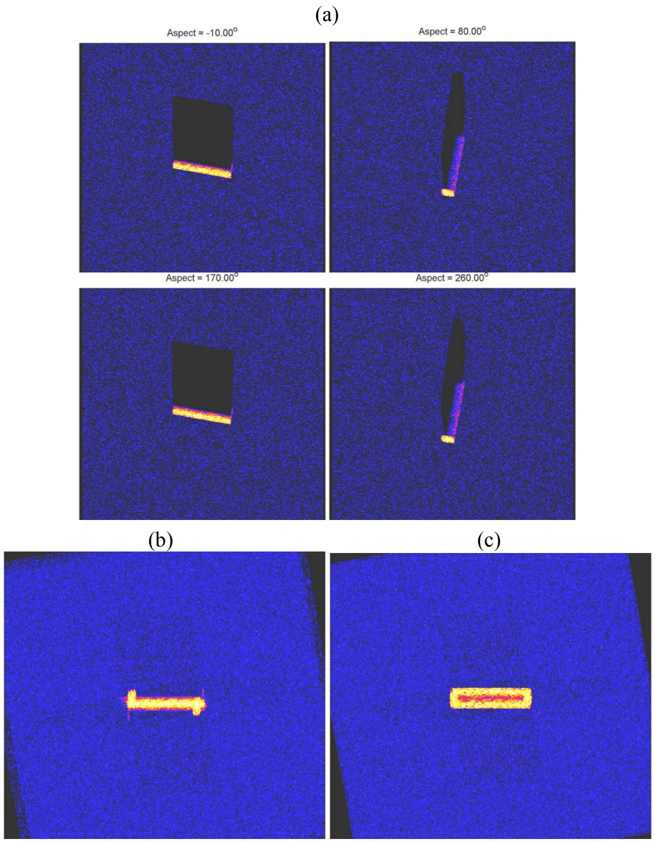

Figure 2. Four simulated views of a cylinder on the seafloor obtained with high-resolution synthetic aperture sonar: (a) single view, (b) registration using correlation technique, (c) correct registration.

The registration itself is not trivial and it can be seen in Fig. 2b that overlaying the four images by using 2D cross-correlation and obtaining the required 2D shift from the peak position in the cross-correlation output does not result in the desired result. Such a method attempts to match highlights in two images, which is not a good criterion, since different parts of the object are viewed (intentionally). A more advanced method to cope with the aspect dependent response is to estimate the target shape (in 3D) from each single view using shape-from-shading techniques. These shapes are not necessarily accurate but they do allow incorporating directionality in addition to intensity for each target pixel. When correlation is then applied, only the overlapping information is used, i.e. pixels that correspond to facets oriented in the same way in both single-view images. This method is explained in more detail in the following section where due to the real data it is a necessary step. In this section it is assumed that the registration can be done, with the adequate registration that results, for this example, in Fig. 2c.

B. Image fusion

After the single-view images are registered successfully, their pixel values can be combined. In this section two approaches are described, applied and evaluated.

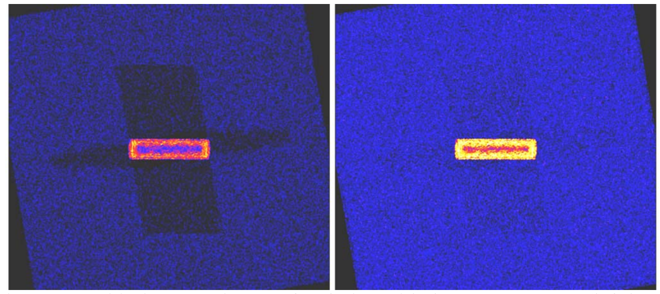

Figure 3. Multi-view SAS image after fusion with the mean operator (left-hand side) versus one after fusion with the max operator (right-hand side).

The first fusion method is based on the mean operator, i.e. each pixel is the average of the single-view pixels. The result for this fusion applied to the standard simulated example in this article is shown on the left-hand side of Fig. 3. For the second fusion method, the mean is replaced by the maximum operator, the result of which is shown on the right-hand side of Fig 3. One could apply several other techniques, but since the maximum operator makes sure that strong highlights in each single-view image are preserved, this approach is preferred. This latter fusion approach also suppresses the shadows much quicker than the former, which is desirable because the shadows are not stationary. Therefore, it is better if the classification is based on merely highlight information. Furthermore, information obtained from shadows has been shown to be strongly dependent on target range and sea floor [5], a variability that is reduced with this approach.

C. Model-based classification

Based on the successful classification results in [5] and [6], the classification features are obtained by matching a template with the snippet. The templates are generated with a validated sonar model [7] for which the mine shape, the geometry and sonar settings are input. The model generates noise-free images that represent what would be expected from a mine. The better the template of a mine matches the snippet, the more probable it is that it is an actual mine. It was shown in [5] that templates are required for each mine type one expects to encounter. Furthermore, asymmetric mines require a set of rotated versions of the target. Variability of other parameters such as bottom slope, bottom type, target tilt or burial was shown to have a much lower impact [5]. In the case of the cylinder classifier, the template bank consists of 16 cylinders rotated at different orientations. The model matching is achieved via cross-correlation of template and snippet, which results in one correlation coefficient (the peak) per template. In Fig. 4 the example of the cylinder is re-visited in an attempt to discriminate it from a rock.

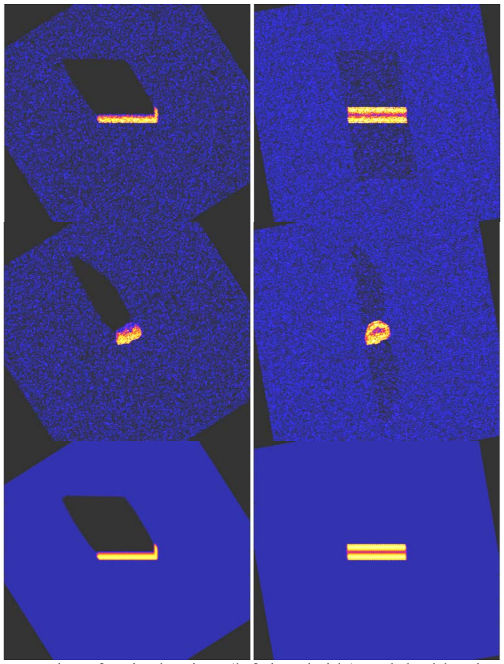

Figure 4. Examples of a single-view (left-hand side) and double-view (right-hand side) of a cylinder (top row), of a mine-sized rock (middle row) and of the template for a cylinder (bottom row).

The top row of Fig. 4 shows two examples of SAS images of the cylinder with, on the left, the single-view SAS image, and on the right, the double-view SAS image. The bank of aspect-dependent templates is generated and correlated with the snippets. The maximum correlation coefficient over all the templates obviously corresponds to the correct aspect template, i.e. the images on the bottom row of Fig. 4. For the single-view snippet the maximum correlation coefficient is 0.64 and for the double-view snippet it is 0.70. The correlation is quite a bit lower than one, which is caused by the fact that the snippet has noisy pixel values (simulated by a Rayleigh distribution here). If the snippet with the rock, a typical representative for a false alarm, is matched with the cylinder templates, the correlation coefficients are much lower: 0.29 for the single-view and 0.27 for the double view. This first example is promising and appears to achieve the objective, which is to obtain good correlation for the correct target, and poor correlation for the false alarm. In this case the separation improves with a second view. Nevertheless, this one example might be misleading. In the following section the model-based correlation will be applied to a large set of these images to evaluate if the correlation coefficient, a strong classification feature, becomes more discriminative. The performance gain of a classifier based on this feature can be evaluated by regarding the statistics of false alarms and the statistics of the targets. In other words, one can generate a receiver operating characteristics (ROC) curve by varying the threshold and counting false alarms and correct classifications.

III. Application to simulated data

In this section the model-based classification is applied to a large dataset. Two objects are investigated, the cylinder representing a mine and the rock representing a typical false alarm. Images are obtained for different ranges and orientations. The parameters mentioned in the previous section are kept and the ranges are varied between 40 and 150 m in 12 steps. The 16 different orientations used are equi-spaced between 0 and 360 degrees anticipating the multiple views, which will be evaluated for 1, 2, 3, 4, 8 and 16 views. In this way the cylinder dataset contains rather similar snippets due to the symmetry, which is not the case for the rock.

D. Model-based correlation coefficient as classification feature.

The correlation coefficient is calculated for each snippet-template pair. This means the result is a set of 16 (number of snippets) times 16 (number of aspect dependent templates) times 12 (number of ranges) cross-correlations for both cylinder and rock. The total number of correlation coefficients for this classification analysis is therefore 2 (cylinder, rock) times 6 (1, 2, 3, 4, 8, 16 views), which is a total of 36864. The aim of the analysis is now to show that it becomes less likely that the rock correlation coefficients are higher than the cylinder correlation coefficients when adding views. The maximum over the templates for the cylinder snippets (that obviously corresponds to the correct match) is shown in the top panel of Fig. 5 for the single-view case. The maximum correlation coefficient clearly increases with range, which is predominantly caused by the increased highlight-to-reverberation ratio, or image contrast. In other words, at short range, for very steep grazing angle, the target echo barely exceeds seafloor scatter. If one does not see the target, one cannot classify it. The angle dependence is also clearly visible in the image. When the cylinder is imaged by the sonar from an endfire direction (θ = 90^o or 270^o), correlation is low and correct classification is more difficult.

The bottom panel of Fig. 5 shows the single-view results for both the cylinder and the rock. The solid lines corresponding to the cylinder are an average of the top image and the dashed lines are the maximum correlation coefficients for the rock. It can be seen that for orientation θ = 0^o, this rock matches a cylinder much better than from other directions. It can also be seen that the rock correlates better at longer ranges, which partly (not completely) cancels the fact that the cylinder correlation coefficient increases with range. These curves suggest that one can now evaluate classification performance. The false alarms (rocks) generate correlation coefficients that have a certain statistical spreading. Estimation of this spreading should make it possible to reveal the relation between probability of correct classification and false alarm rate, which is the objective of this article. In the following subsection the analysis will be continued by generating the results of Fig. 5 for the multi-view cases.

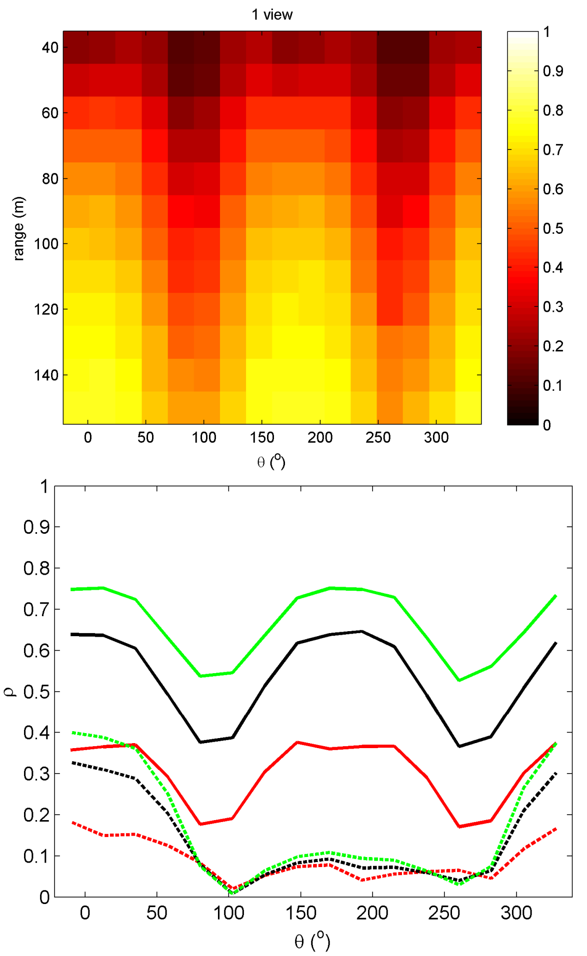

Figure 5. Top: maximum correlation coefficient between cylinder snippet and cylinder templates versus aspect and range (top). Bottom: maximum correlation coefficient between cylinder (solid) and rock snippet (dashed) and cylinder templates. Plots are an average for short range (< 75 m) in red, medium range in black and long range (> 115 m) in green.

E. Performance versus number of views

In the previous section it was explained that the classification performance is primarily determined by the correlation coefficients of the snippet with its corresponding template. Fig. 6 shows the results in exactly the same way as Fig. 5 for 2, 3, 4, 8 and 16 views, respectively. The rationale behind these values is the fact that mine hunting missions are often conducted as efficient lawn-mower patterns that can be repeated in different directions. The views are equi-spaced, thus for instance the four-view case can consist of aspects 20°, 110°, 200° and 290°. The 3-view case is somewhat peculiar, but has been added because it is useful for surveys. One can imagine that the area is covered once with a lawnmower pattern. If the situation turns out to be difficult, the lawn-mower can be repeated in cross-hatching mode with denser spacing such that each seafloor patch is viewed thrice. In general, the situation improves when views are added. The correlation coefficients rise and also become less aspect dependent. Especially the difference between 2 and 3 views is rather significant, which is primarily caused by the fact that the difficult endfire situation does not exist anymore (at least one of the views is always well away from endfire).

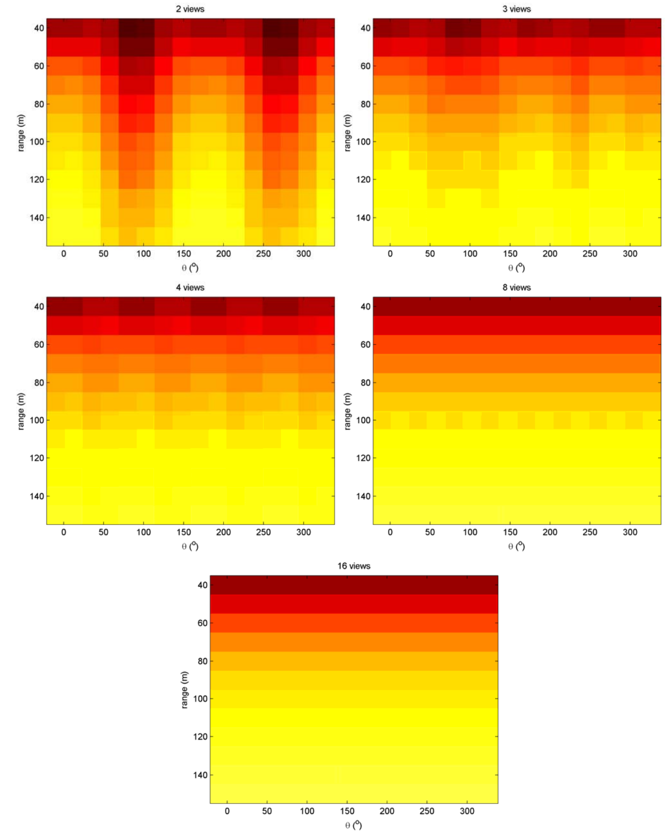

Figure 6. Maximum correlation coefficient between cylinder snippet and cylinder templates versus aspect and range for 2, 3, 4, 8 and 16 views.

Fig. 7 shows the results for both cylinder and rock. The angular variability decreases for more views for both cylinder and rock. The distance between minimum cylinder score and maximum rock score also increases with number of views.

Even if Fig. 7 is a clear sign that it is easier to discriminate a cylinder from a rock when more views are fused, the plots are not yet proof that the classification actually improves. For that, the impact on false alarm rate will need to be investigated, which will be done in Section V. However, first the multi-view fusion is demonstrated on experimental data in the following section.

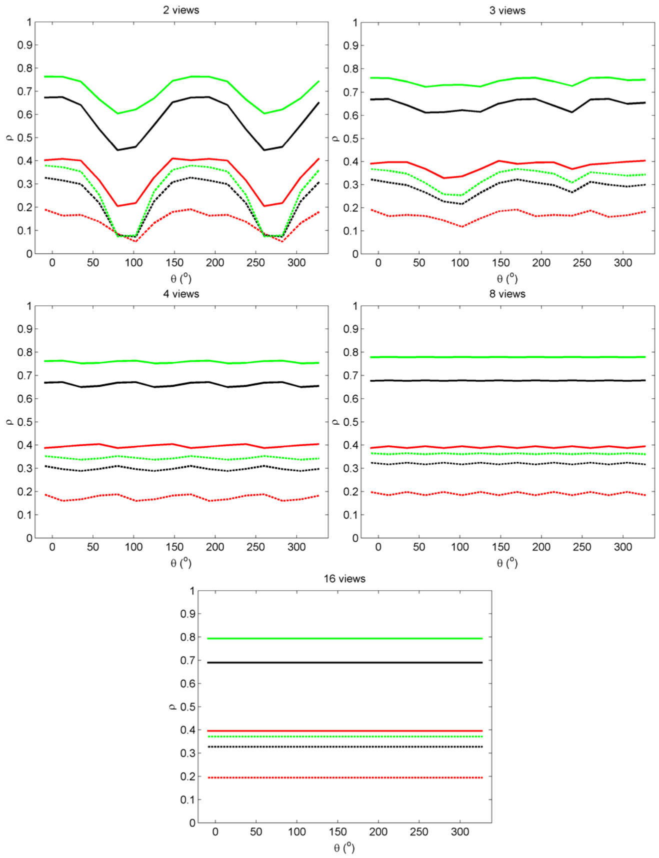

Figure 7. Maximum correlation coefficient between cylinder (solid) and rock snippet (dashed) and cylinder templates. Plots are an average for short range (< 75 m) in red, medium range in black and long range (> 115 m) in green.

IV. Application to Experimental Data

The NATO Undersea Research Centre develops the MUSCLE autonomous underwater vehicle equipped with a SAS to demonstrate the concept of autonomous mine hunting. Images produced by this system typically take the form of rectangular tiles covering areas of 50m by 110m, with a resolution of about 2.5cm. After every mission, these images (or "tiles") are geo-referenced using navigation information and stored in a database. For the target multi-view process, only those tiles that cover the target region are considered.

The set of relevant tiles is first sorted by observation direction. This is convenient for the subsequent matching procedure that will register them in pairs, as it is frequently easier to match a tile to those that have been acquired from a similar orientation.

The tiles are then coarsely matched in order using a naïve registration process that uses normalized cross-correlation of the images. Since target returns are frequently brighter than their surroundings, only those pixels that are above the mean intensity are considered. This procedure is not enough to completely register the different views, but is enough to get them all matched within target-size positioning error. Tile heading errors are assumed to be small enough that they can be discarded. That is, the different views of the target are only mismatched by a translation.

By matching the different views in order, a cumulative registration error can be computed by adding up the consecutive translation vectors that match every pair of views. This can be used as a sanity check for the process, and if the accumulated displacement is about the size of the target we consider the coarse matching successful. Otherwise, one or more views probably suffer from low image quality and should be discarded.

The coarsely matched images are then inverted [8] to obtain a 3D approximation to the target shape and its immediate surroundings, from which the surface normal vectors are calculated and used for the fine registration of the different views. Unlike the images of the targets—which can appear very different with changing highlight and shadow positions—the normal vectors should be independent of the direction of observation, and are therefore useful for the finer matching. The score function for the fine matching is the dot product of the normal maps, which highlights the location that best fits every pair of views.

Once all views are precisely co-registered, they can be combined in several ways. The simplest approach to obtain a combined 2D image is to compute the average of the different views. Other approaches are to compute the median or the max of the backscatter levels for every pixel, which have the advantage of discounting shadowed pixels.

For the derivation of the combined 3D view, the different elevation maps are averaged using their corresponding backscatter images as weights. In this way, pixels in shadowed areas have no influence in the final result, while those corresponding to prominent features are given more relevance.

A. *Limitations*

The automatic fine-registration process assumes all the views are meaningful, but in some conditions spurious image artifacts or degraded image quality can make the automatic process to fail. Sonar images taken in very shallow waters, for instance, will present several echoes and blurring that can prevent the fine-registration to converge.

Errors in the determination of sensor heading are also occasionally big enough that matching target views will require a transformation more complicated than a single translation. In general, it is convenient to have the mission navigation filtered and corrected before attempting to produce multi-view reconstructions.

Finally, the inversion process assumes the observed scene can be represented by a single valued elevation surface [4]. For some object shapes this may be too simplistic and could lead to small errors in the last fine-registration stage, which may nevertheless produce a meaningful final result (given that the median and weighted averages, will discard most of the outliers).

In general, most of these limitations will disappear by taking a top-down approach to the fine-registration process. This would start by processing the whole mission making sure navigation errors are corrected as early as possible. The use of SLAM techniques fits well with multi-pass missions and greatly helps in reducing the biggest overall orientation and translation errors. After that, 3D inversion of every tile would be used for its rectification; this would compensate for differences in seafloor elevation that can cause projection distortions that could affect the fine matching.

Furthermore, a top-down approach greatly constrains the uncertainty in position of the different views of an object. This means that in the final normal matching stage the potential differences in position and orientation would be so small that the process will very likely succeed. For this reason, our current work is oriented to the production of multi-scale mosaics of the mission area that are not only useful for the target multi-view matching, but also to produce precise wide area maps of the surveyed regions.

B. Cylinder

Fig. 8 shows the results of the multi-view fusion applied to a cylindrical object. The four images in the top row are some of the individual views of the object, and show a clear dependence to the direction of observation, producing acute variations in highlight and shadow positions. The result of the multi-view fusion, however, removes the influence of all those transitory effects and shows a clear picture of the object's geometry.

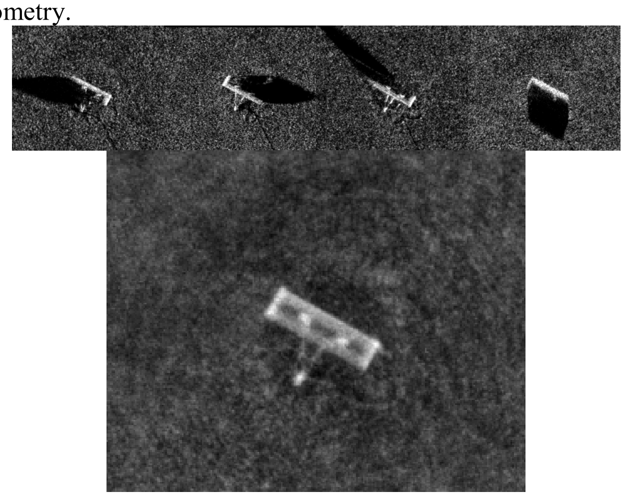

Figure 8. Fusion of several views of a cylindrical target. The top row shows 4 of the 15 independent views. Main image presents the fusion result after fine co-registration.

C. Wedge-shape

Fig. 9 shows multi-view reconstruction results for a wedge object. For the ATR purposes discussed in this paper, the resulting combined 2D image is the most relevant, but the additional 3D model obtained by the multi-view fusion procedure could also be used for more advanced target identification methods or for direct validation by a human operator.

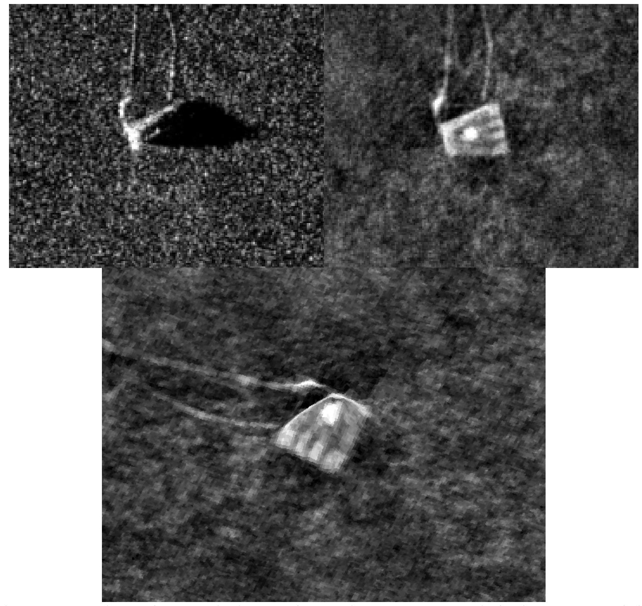

Figure 9. Fusion of several views of a wedge target. Top left shows one of the less favorable single views. Top right shows the result of the multi-view fusion, offering a clear view of the object. Main image shows the 3D reconstruction of the target, which is another output of the fine registration process.

V. FALSE ALARM RATE REDUCTION

The analysis conducted in Section III indicated that the cylinder correlation coefficients become higher and less aspect dependent. It was also mentioned that this indication is no proof of enhanced classification, because the behavior for the rock (= false alarm) correlation coefficient should be taken into account as well. In this section the link with the false alarm correlation coefficient statistics is established based on a simple classification approach and an interesting observation on the false alarm statistics.

A simple approach to discriminate between the cylinders and the rocks is to apply a threshold to the maximum correlation coefficients of Section III and classify each snippet that exceeds the threshold as a cylinder. The threshold ρ is now set as high as possible, but such that no cylinders are missed. The threshold can be obtained from the images in Fig. 5 and Fig. 6 and for instance would be 0.40, 0.48, 0.63, 0.67, 0.70, 0.72 for the different number of views at a range of 100 m. The observation that was now made was that the distribution of the correlation coefficients for the rock does not vary significantly as a function of the number of views. The mean and standard deviation for the total set of cylinders (μ_{cylinder},σ_{cylinder}) and for the total set of rocks (μ_{rock},σ_{rock}) is given in Table 1.

TABLE I
MEAN AND STANDARD DEVIATION OF MAXIMUM CORRELATION COEFFICIENTS FOR THE DIFFERENT VIEWS.

Number of views	1	2	3	4	8	16
$\mu_{cylinder}$	0.50	0.54	0.59	0.60	0.62	0.63
$\sigma_{cylinder}$	0.09	0.08	0.02	0.01	0.00	0.00
μ_{rock}	0.20	0.22	0.23	0.24	0.25	0.25
σ_{rock}	0.08	0.08	0.07	0.06	0.07	0.07

The values in Table 1 indicate that the performance of a threshold-based classifier as described above, indeed is mainly affected by the correlation coefficients of the target (cylinder). Under the assumption that the false alarm values behave like a Gaussian distribution with average μ_{rock} and standard deviation σ_{rock}, the probability that a false alarm exceeds the threshold can be calculated with:

$$P_{fa} = \frac{1}{\sigma_{\text{rock}}\sqrt{2\pi}} \int_{\rho}^{\infty} e^{-\frac{(x-\mu_{\text{rock}})^2}{2\sigma_{rock}^2}} \, dx \qquad (2)$$

When assuming that the false alarm correlation values have a Gaussian distribution, one runs into the problem that a correlation can never exceed one, whereas the Gaussian distribution can. However, since one is always many (>9) standard deviations away from the mean μ_{rock}, this discrepancy has a negligible effect on the probability P_{fa}, i.e. it is below machine precision for $\rho>1$.

Fig. 9 shows the results based on this statistical model. On the left-hand side the probability of false alarm is plotted versus range for the six multi-view cases. It is clear that adding the third view is a major improvement. On the right-hand side only the target range of 100 m is shown for the different multi-view cases. The false alarm reduction is enormous, more than two orders of magnitude for both addition of the third and fourth view. For example when a million false alarms are generated, typically only 95380 are left after the second view and just two false alarms when incorporating the third view. It can also be seen that beyond the fourth view there is not much to gain.

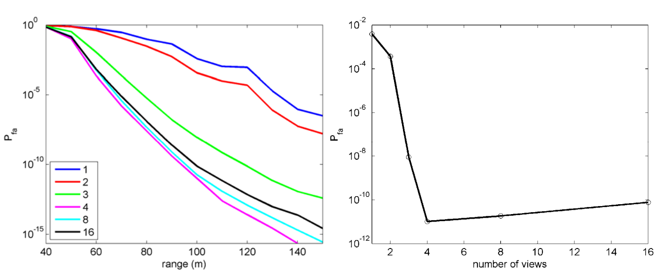

Figure 9. Probability of false alarm for the different number of views as a function of target range (left-hand side) and probability of false alarm for a fixed target range of 100 m (right-hand side).

The results are promising, and multi-view SAS imaging should be regarded as the solution for MCM in high-clutter areas. It is known that the multi-view fusion is more difficult in such environment, primarily because target snippets may be associated incorrectly when there are so many of them. One should also note that the data in this analysis are limited. The rock, even if many of rotated versions were used, only represents a subset of the false alarms one may encounter.

VI. CONCLUSION

This article presents a sea mine classification approach for multi-view synthetic aperture sonar image fusion and evaluates the value of it based on a realistic simulation case. It is shown that classification performance improves significantly when adding the second, third and fourth view. The value of additional views beyond four is rather limited. The technique is also shown on two real data examples.

Another study [9] has shown that target classification performance worsens as the complexity of the seabed increases because of the false alarms generated by seabed features. Additionally, when the seabed is characterized by sand ripples, the appearance of the ripples in the resulting imagery – and the relative difficulty of performing detection and classification -- exhibits a strong dependence on the orientation of the data-collection survey route [10]. Therefore, future work will assess the merits of the proposed multi-view classification approach on data containing such complex seabeds.

REFERENCES

[1] D. Williams and J. Groen, "Multi-View Target Classification in Synthetic Aperture Sonar Imagery," *Proceedings of the 3rd International Conference & Exhibition on Underwater Acoustic Measurements (UAM)*, Nafplion, Greece, June 2009, pp. 699-704.

[2] V. Myers and D. Williams, "A POMDP for Multiview Target Classification with an Autonomous Underwater Vehicle," submitted, 2010.

[3] V. Vapnik, Statistical Learning Theory, Wiley, 1998.

[4] E. Coiras, Y. Petillot, D. M. Lane, "Multi-Resolution 3D Reconstruction from Side-Scan Sonar Images", *IEEE Transactions on Image Processing*, Vol. 16, No. 2, pp. 382-390, February 2007.

[5] J. Groen, E. Coiras, J. Del Rio Vera, B. Evans, "Model-based sea mine classification with synthetic aperture sonar", *IET Journal of Radar, Sonar and Navigation*, Vol. 4, No. 1, pp. 62-73, 2010.

[6] H. Midelfart, J. Groen and Ø. Midtgaard, "Template matching methods for object classification in synthetic aperture sonar images", *Proceedings of the 3rd International Conference & Exhibition on Underwater Acoustic Measurements (UAM)*, Nafplion, Greece, June 2009.

[7] E. Coiras, A. Ramirez-Montesinos, J. Groen, "GPU-based Simulation of Side-looking Sonar Images", *Proceedings of the Oceans 2009 conference*, Bremen, Germany, May 2009.

[8] E. Coiras, J. Groen, "3D target shape from SAS images based on a deformable mesh", *Proceedings of the 3rd International Conference & Exhibition on Underwater Acoustic Measurements (UAM)*, Nafplion, Greece, June 2009.

[9] J. Groen, E. Coiras, and D. Williams, "Detection Rate Statistics in Synthetic Aperture Sonar Images," *Proceedings of the 3rd International Conference & Exhibition on Underwater Acoustic Measurements (UAM)*, Nafplion, Greece, June 2009, pp. 367-374.

[10] D. Williams, "Bayesian Data Fusion of Multiview Synthetic Aperture Sonar Imagery for Seabed Classification," *IEEE Transactions on Image Processing*, Vol. 18, No. 6, pp. 1239-1254, June 2009.