ATR FUSION IN SAS IMAGERY FROM THE MUSCLE SONAR ON COMPLEX SEABEDS

JD Woodward Dstl, UK
GL Davies Dstl, UK

1 INTRODUCTION

The purpose of this paper is to understand the different false alarm densities produced by manual analysis of SAS (Synthetic Aperture Sonar) imagery and different algorithm approaches undertaking a mine-detection task and to illustrate the benefits of fusion for reducing false alarm density. Understanding the difference in detection probability and false alarm densities between ATR (Automatic Target Recognition) algorithms and manual analysis is important for a number of reasons. Firstly, it allows informed decisions to be made on how much support ATR can provide to operators in different environments. Secondly it provides a benchmark against which improvements in false alarm density can be measured.

In this paper probabilities of detection and false alarm densities produced by four ATRs developed by universities in the UK are compared with students performing manual analysis in the task of detecting mine-like objects in SAS imagery. Different fusion schemes are explored to reject false alarms in highly cluttered environments. Receiver Operating Characteristics (ROC) curves are used to show the trade-off between probability of detection and the number of false alarms.

False alarms are generated in complex areas by clutter such as stones and ripples which may appear mine-like from a given aspect. High false alarm densities can be generated when imaging complex areas from multiple angles as there is an increased chance of viewing the clutter objects from an aspect at which they appear mine-like. On areas of complex seabed minehunting is an unbalanced problem with the number of mines outweighed by the number of clutter objects.

Several post-processing schemes have been proposed to remove false alarms and improve probability of detection. Two techniques are considered in this paper: multiple ATR fusion[1-6] in which contacts from several ATRs are fused, and multiple aspect fusion in which contacts corresponding to different views of the same geographic position are fused. Simple fusion schemes include voting (M out of N ATRs generated a contact for this object) whilst more sophisticated techniques may fuse ATRs based upon prior knowledge of their performance, for example by using the Dempster-Shafer theory of evidence[2].

Multi-aspect fusion takes advantage of the fact that the similarity of seabed clutter to a mine can change depending on its orientation. Fusing the information from multiple aspects has been shown to provide a significant reduction in the number of false alarms[1]. One example commonly used to illustrate why this works is the case of a cylindrical mine compared with a spherical rock. When viewed end-on, within SAS imaging, the rock and cylinder appear very similar, but when the viewing angle changes by 90 degrees the shape of the cylinder will have changed significantly whereas the rock still looks like the end of the cylinder. Hence by combining the information across multiple aspects the false alarms which appear mine-like only from certain aspects can be removed, leaving only those objects which appear genuinely mine-like from all aspects.

1.1 Dataset

The dataset used for evaluation consists of SAS data from the NATO Centre for Maritime Research and Experimentation (CMRE) Minehunting UUV for Shallow water Covert Littoral Expeditions (MUSCLE) system collected during the COLOSSUS 2 sea trial. The advantages of using this dataset for evaluation are:

- it covers three seabed types of varying difficulty for minehunting: flat sand (easy); mud and stones (moderate); and boulders and sand ripples (hard);
- the high resolution (2.5 cm x 1.5 cm) MUSCLE SAS data is most likely to be representative of the quality of sensor data available from future minehunting systems.

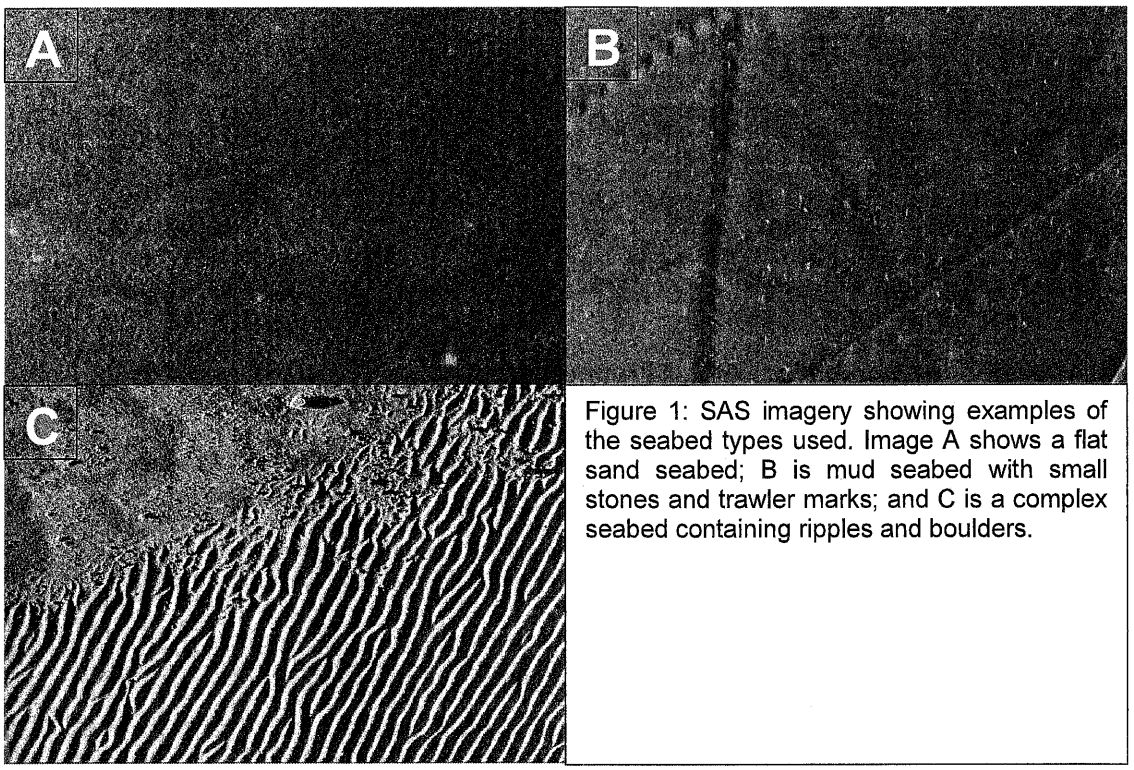

Figure 1: SAS imagery showing examples of the seabed types used. Image A shows a flat sand seabed; B is mud seabed with small stones and trawler marks; and C is a complex seabed containing ripples and boulders.

Four target types were deployed in these areas: a truncated-cone, a cylinder, wedge and a calibrated rock. Each area contained nine targets imaged on average six times from a range of different aspects.

1.2 Manual Analysis

Dstl has carried out manual analysis using two university students. The students were initially trained on a sub-set of the data from the COLOSSUS 2 sea trial reserved for training ATR algorithms and then evaluated on the standard evaluation dataset for each area. Training consisted of presenting a series of SAS images containing targets to the students and asking the students to mark the position of all targets in the image; the actual target positions were then indicated allowing the students to learn what the targets look like in SAS images. The students were required to provide a confidence value ranging from from 1 to 5 for each detection; this was used to generate ROC curves.

2 OBSERVATIONS ON THE DIFFERENCE BETWEEN RECEIVER OPERATOR CHARACTERISTIC CURVES FROM ATR AND MANUAL ANALYSIS

The typical output from different algorithmic approaches in each area is shown by plotting a ROC curve for each ATR; these show Pd (the ratio of the number of detections to the number of detection opportunities) plotted against the number of false alarms per square nautical mile. The points on the ROC curves are generated by varying the threshold input parameter to the ATR. In general, the Pd and the number of false alarms both increase as the ATR threshold parameter is reduced: the optimum threshold value to use is a compromise between the required value of Pd and an acceptable number of false alarms.

Figure 2 shows the typical output of each ATR in an area of flat sand, a benign environment in which targets visually stand out clearly against the seabed. All achieve a maximum Pd between 0.9 and 1.0 and the number of false alarms per square nautical mile varies between zero (for ATR 3) to 500 for ATR 2. Manual analysis produces ROC curves that are very similar to ATR 1.

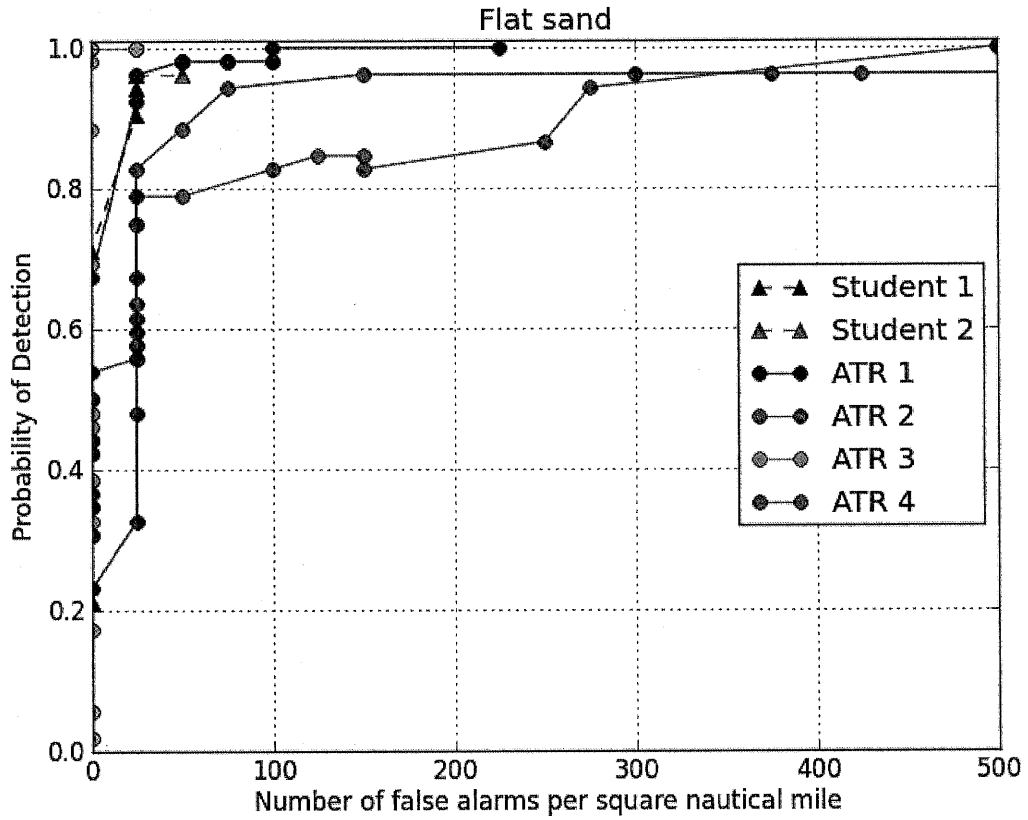

Figure 2: Comparison of ROC curves produced by ATR algorithms with manual analysis against a flat sand seabed.

Figure 3 shows the ROC curves for each ATR in an area of soft mud with stones and trawler marks. This is a moderate environment in which targets generally stand out well from the seabed but the seabed features and stones cause the ATRs to produce a higher number of false alarms. All ATRs achieve a maximum Pd between 0.9 and 1.0, as they did in the flat sand area, but the number of false alarms per square nautical mile has increased to between 80 and 500. The manual analysis

produced a similar ROC curve to ATR 2, which lies somewhere between the best and worst performing ATR algorithms.

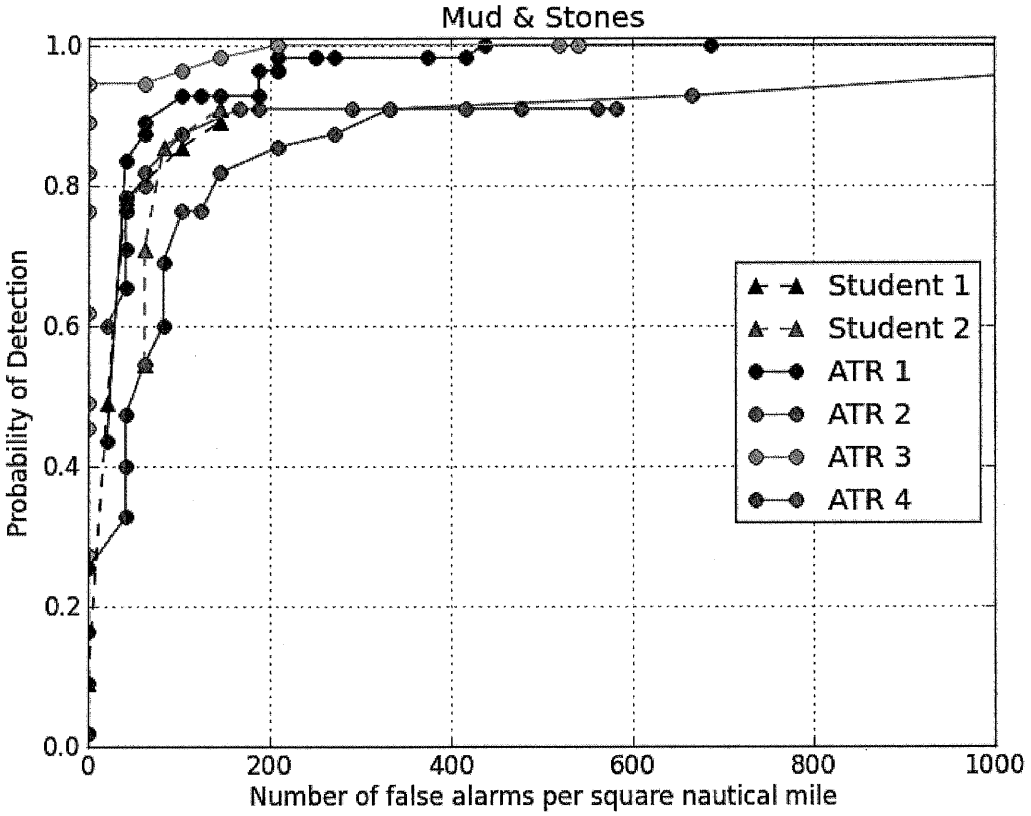

Figure 3: Comparison of ROC curves produced by ATR algorithms and manual analysis against a soft muddy seabed with stones.

Figure 4 shows the ATR ROC curves in a difficult environment. The area contains a mixture of seabed types of varying difficulty including areas of sand ripples and boulder fields as well as some patches of flat seabed. The ATRs typically reach a maximum Pd of approximately 0.9 with over 5000 false alarms per square nautical mile. The number of false alarms produced by ATRs is approximately five to ten times greater than the number of false alarms produced by manual analysis in this environment.

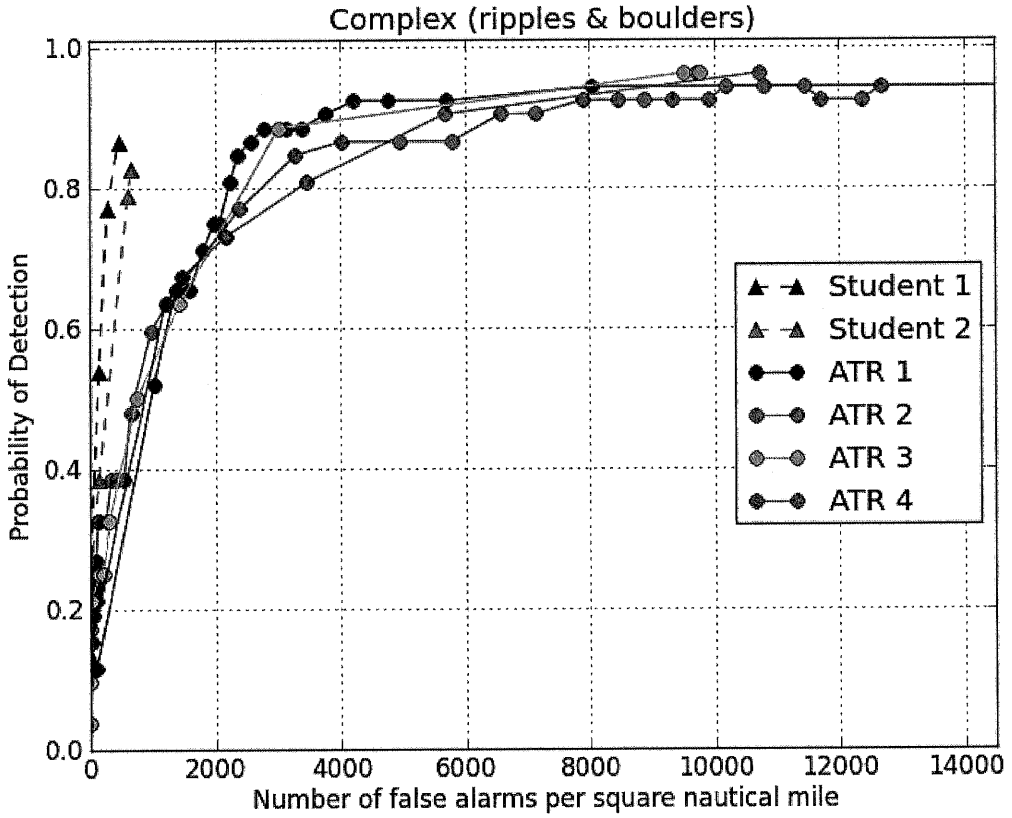

Figure 4: Comparison of ROC curves produced by ATR algorithms and manual analysis against a seabed containing sand ripples and boulders.

The results presented in this section of the report have shown that ATRs can achieve comparable probabilities of detection and false alarm rates to manual analysis in benign conditions but that in more complex environments, ATRs produce far more false alarms than manual analysis (by a factor of between five and ten).

3 MULTI-ATR FUSION

In this section we investigate the benefit of multiple algorithm fusion using ATRs that have a similar performance but use different algorithmic approaches. Fusing multiple ATR outputs using voting methods has previously been shown to produce a measurable reduction in false alarms[13]. Simple voting methods include a logical OR, unanimous consensus (logical AND), and majority voting.

3.1 Methodology

In order to carry out multi-ATR fusion, it is necessary to group contacts together into clusters where contacts within a cluster all lie close to the same physical location (an association distance is specified to take into account navigation errors and ATR marking errors). A rule is then applied to the contacts within a cluster to decide whether to call a detection (output a fused contact corresponding to the cluster) or discard the cluster.

3.2 Results

Figure 5 shows the results of fusing five ATR algorithms using majority voting in the form of a ROC curve; this shows that the performance of the fused ATRs is better than any individual ATR. Table 1 lists the false alarm density (number of false alarms per square nautical mile) corresponding to each individual ATR at a Pd of 0.8. A Pd of 0.8 was chosen for comparison as it is close to the highest common Pd achieved by both the algorithmic and manual analysis methods. The false alarm density of the fused result is reduced by 18% compared to the best performing individual ATR.

Name	False alarm Density at Pd=0.8	Number of False alarms generated by each ATR relative to ATR 1 at Pd=0.8
Student 1	310	14%
Student 2	610	27%
Majority	1810	82%
ATR 1	2200	100%
ATR 3	2480	112%
ATR 4	2730	124%
ATR 2	3320	150%

Table 1: False alarm density (per square nautical mile) corresponding to a Pd of 0.8.

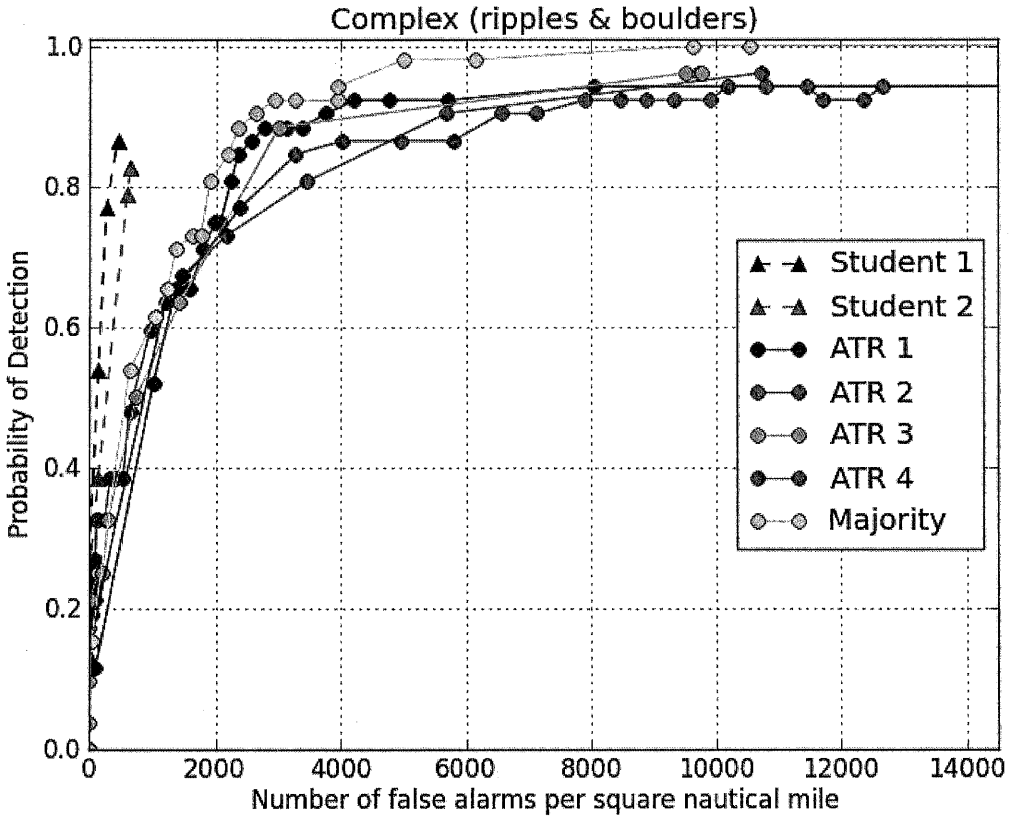

Figure 5: ROC curves for manual analysis, individual ATR algorithms, and majority voting fusion (shown by yellow line with circles) on the complex seabed.

Figure 6 compares the results of fusing ATR contacts using different rules: it suggests that the rules produce similar results, although majority voting provides slightly better performance than the rest for this dataset.

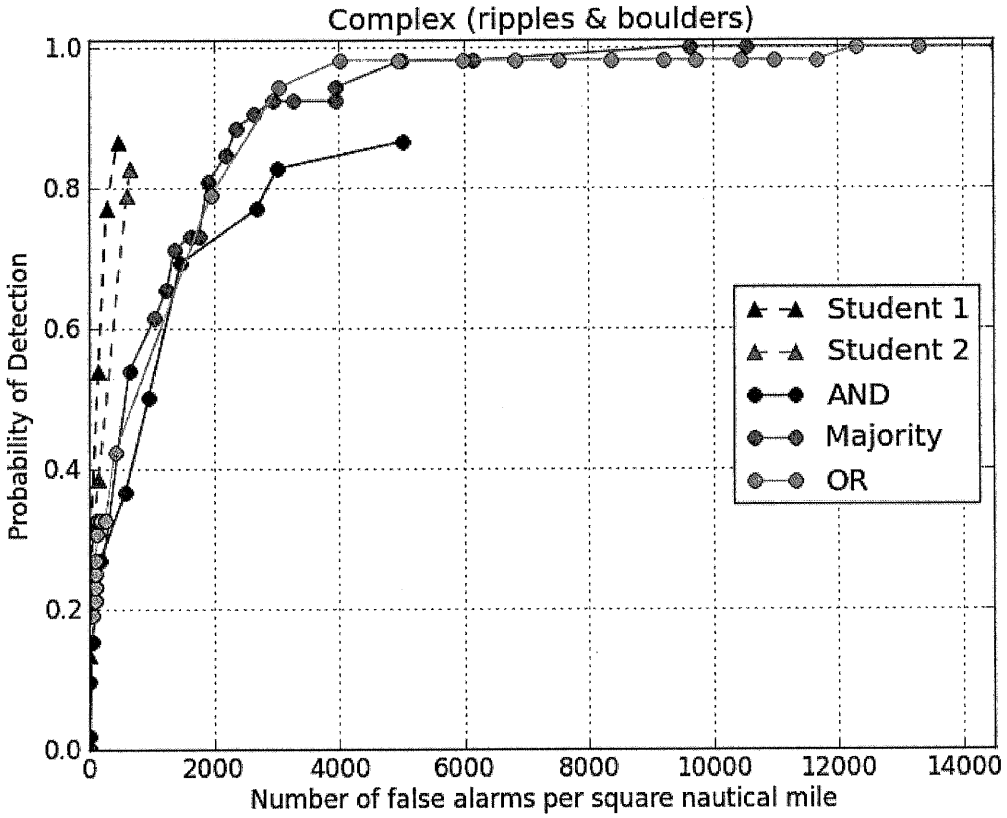

Figure 6: Comparison of simple voting schemes for fusion of the four ATRs on a complex seabed.

4 MULTI-VIEW FUSION

4.1 Introduction

Multi-view fusion is the process of fusing ATR contacts produced from different views of the same physical object. In order to analyse ATR performance resulting from multi-view fusion it is necessary to alter the definition of Pd that has been used previously in this paper. Instead of defining Pd as the ratio of the number of ATR contacts corresponding to correct detections to the total number of detection opportunities, the definition is changed to the fraction of targets (mines) detected (this is also the definition of Pd or, using NATO nomenclature, Bd, that is used operationally). The disadvantage of using this definition of Pd is that there are usually fewer physical targets than detection opportunities so that the confidence intervals on estimates of Pd are larger and the values of Pd are more dependent on the mission plan (a mission plan that results in many detection opportunities per target is likely to allow an ATR or human conducting manual analysis to achieve a higher Pd for each target than one that only results in a single detection opportunity per target). Treating the Pd as the fraction of detected targets will generally lead to higher values of Pd than one based on detection opportunities: for example, if a target is detected on 2 out of 4 detection opportunities, the fraction of detected targets value of Pd is 1.0, whereas the per detection opportunity value of Pd is 0.5. In order to carry out multi-view fusion, contact clusters must be created using a similar methodology to that described in Section 3.1.Errors in contact positions due to the navigation system and ATR marking errors need to be taken into account when forming the clusters. When repeatedly surveying an object the standard deviation of the marking positions is 1.4

meters. No attempt has been made to correct for these navigational errors by co-registration or similar processes.

4.2 Results

4.2.1 N-Looks

For the N-looks rule a detection is called if there are at least N contacts (from different images) in a cluster of contacts. Figure 7 shows results obtained by fusing contacts corresponding to multiple views using the N-looks rule; the average number of looks at a target is six. For values of N of 2 and 3 the false alarm density is reduced by about two thirds without affecting the Pd; however, for N $\geq$ 4 the Pd degrades. These results suggest that (in this case, at least) N needs to be less than or equal to half the average number of looks to avoid degrading Pd.

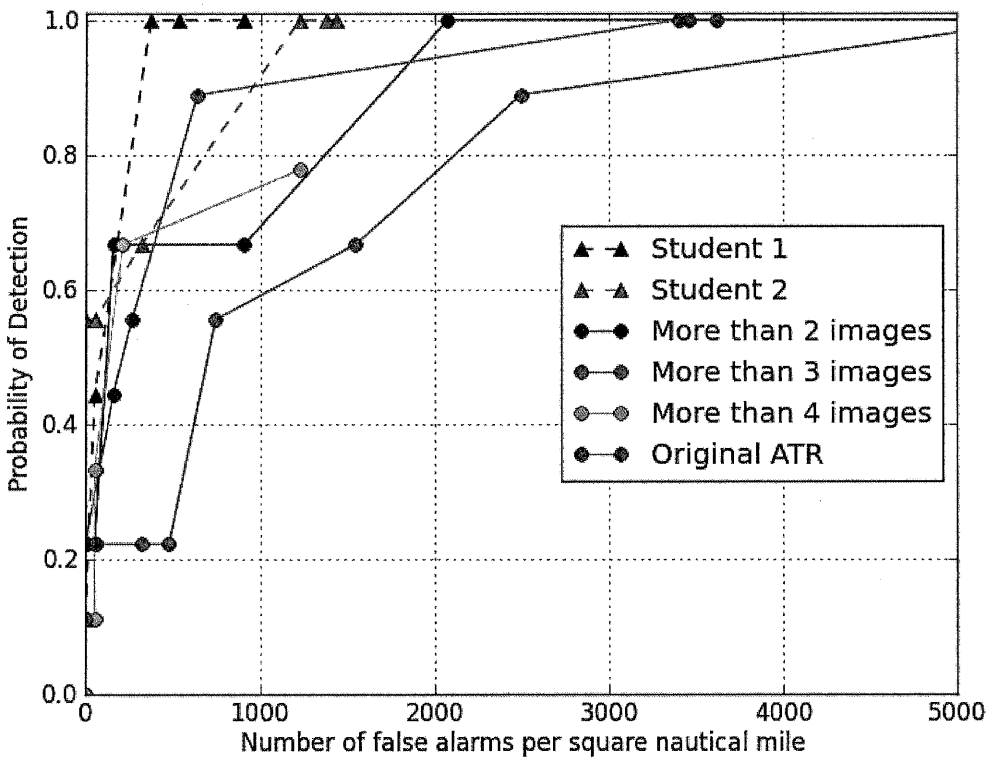

Figure 7: Result of fusing contacts for multiple views by applying the N-looks rule to ATR 3 on the complex seabed. Manual analysis curves for the students are shown for comparison (but these have not been fused).

4.2.2 Using Confidences

The results described in Section 4.2.1 ignored the confidence value that is returned by some of the ATR algorithms for each contact; contacts output by the ATR for each threshold value were treated with equal weight in the fusion process. In this section, results of multi-view fusion taking into account confidence values are presented for an ATR. The fusion process consists of two stages: firstly, contacts in a cluster that correspond to the same aspect (≤5° aspect change) are fused, the fused confidence value is given by the maximum of the confidence values corresponding to the contacts being fused (detections are assumed to be dependent); secondly, contacts from different aspects (≥5° aspect change) are fused, the fused confidence value is given by the below equation (the detections are assumed to be independent).

$$c_F = 1 - \prod_i^N (1 - c_i)$$

Where c_i is the confidence value associated with each aspect; c_F is the fused confidence value.

The ROC curves resulting from this fusion process were produced by thresholding the fused contacts by confidence value. Figure 8 shows the results of fusing contacts produced by the ATR 3 for the complex seabed taking into account confidence values. It can be seen that multi-aspect fusion in this case has resulted in similar performance to that of manual analysis. It should be noted, however, that the ROC curves corresponding to manual analysis have not undergone any fusion processing.

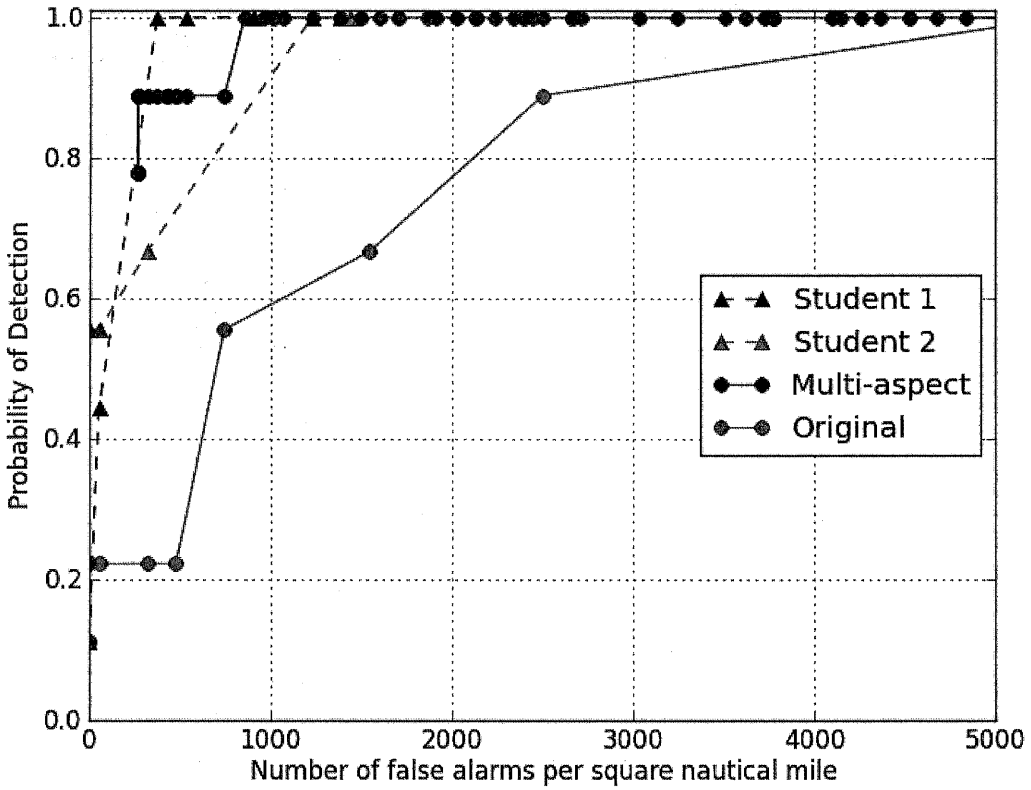

Figure 8: Multi-aspect fusion applied to ATR 3 on the complex seabed. Manual analysis ROC curves are shown for comparison (but these have not been fused). Note the clustering algorithm introduces some uncertainty into the form of the ROC curve.

5 DISCUSSION & CONCLUSIONS

In a benign environment the ATRs from UK universities were observed to perform as well or better than the manual analysis conducted by students. In a complex environment containing ripples and boulders the ATRs generated considerably more false alarms than manual analysis (on the order of 4-5 times as many). Hence whilst the ATRs could be deemed to already be on a par with manual analysis in benign environments additional post-processing must be done to reject enough false alarms to allow them to function usefully in complex and cluttered environments.

Fusing the outputs of multiple ATRs using voting yielded an 18% reduction in the number of false alarms generated. Little difference was observed between different "M of N" voting schemes with a majority vote providing marginally better results. It would be of interest to see if a more sophisticated fusion scheme with knowledge of the relative performance of the ATRs could be used to further reduce false alarms. One method would be to implement the Dempster-Shafer fusion scheme used by Sawas in sidescan sonar imagery[14] on the complex seabed used here.

Multiple aspect fusion was shown to produce the greatest reduction in false alarms for a cluttered environment. For ATR 3 multi-aspect fusion showed an eight-fold reduction in false alarms (from 2000 to 250 per square nautical mile) whilst maintaining a 0.8 detection probability. This reduced the false alarm density of the ATR to the level of manual analysis at a single look. It should be noted that the results from multiple aspect fusion are based on the detection of nine targets in a complex region with on average 6 views per targets. This is in contrast to the much larger dataset available for single-look-detection where 56 detection opportunities are available. These results are robust to navigational error of the MUSCLE AUV as no attempt was made to correct for this using co-registration of images or otherwise.

The following conclusions can be drawn from the analysis that has been carried out on the performance of ATR algorithms from the academic literature using the SAS data from the COLOSSUS 2 sea trial:

- In benign environments an algorithmic approach produces similar false alarm densities and detection probabilities as manual analysis;
- the false alarm density produced by ATRs on a complex seabed can be five to ten times that of manual analysis;
- fusing the outputs from multiple ATRs can reduce the number of false alarms (false alarms were reduced by 18% following the fusion presented in this paper) on a complex seabed;
- fusing contacts produced by a single ATR corresponding to multiple views can reduce the false alarm density on a complex seabed to a similar level to that of manual analysis (note that in this case does not take into account any fusion of contacts produced by manual analysis).

It should be remembered that these conclusions are based on the analysis of a dataset containing high quality, high resolution SAS imagery collected by one sonar in three areas containing different seabed types. Further analysis of data collected in different environments and by different sonars would be needed to understand whether these conclusions apply more generally.

6 ACKNOWLEDGEMENT

This work has been carried out in the Platform Sciences Group of Dstl Physical Sciences. The data used in this work was provided by the NATO Centre for Maritime Research and Experimentation (CMRE). This paper contains British Crown copyright material that is published with the permission of the Controller of Her Majesty's Stationery Office.

7 REFERENCES

1. J. Groen, E. Coiras and D. P. Williams, presented at the OCEANS 2010 IEEE - Sydney, 2010 (unpublished).
2. P. Y. Mignotte, E. Coiras, H. Rohou, Y. Petillot, J. Bell and K. Lebart, IET Radar Sonar Navig. **2** (2), 146-154 (2008).
3. D. L. Hall and J. Llinas, Proceedings of the IEEE **85** (1), 6-23 (1997).
4. C. M. Ciany and W. Zurawski, presented at the OCEANS '02 MTS/IEEE, 2002 (unpublished).
5. D. Clark, S. Julier, R. Mahler and B. Ristic, presented at the Sensor Signal Processing for Defence (SSPD 2010), 2010 (unpublished).
6. T. Aridgides, M. Fernandez and G. Dobeck, in *Detection and Remediation Technologies for Mines and Minelike Targets Vi, Pts 1 and 2*, edited by A. C. Dubey, J. F. Harvey, J. T.

Broach and V. George (Spie-Int Soc Optical Engineering, Bellingham, 2001), Vol. 4394, pp. 1123-1134.

7. B. F. J. H. D. Zerr, in *Underwater Acoustic Measurements : Technologies and results* (Nafplion, Greece, 2009).
8. V. Myers and D. P. Williams, IEEE J. Ocean. Eng. **37** (1), 45-55 (2012).
9. M. R. Azimi-Sadjadi, A. A. Jamshidi and G. J. Dobeck, in *Detection and Remediation Technologies for Mines and Minelike Targets Vi, Pts 1 and 2*, edited by A. C. Dubey, J. F. Harvey, J. T. Broach and V. George (Spie-Int Soc Optical Engineering, Bellingham, 2001), Vol. 4394, pp. 1173-1182.
10. J. Fawcett, V. Myers, D. Hopkin, A. Crawford, M. Couillard and B. Zerr, IEEE J. Ocean. Eng. **35** (4), 863-876 (2010).
11. J. G. David Williams, Proceedings of the 3rd International Conference and Exhibition on Underwater Acoustic Measurements: Technologies and Results (2009).
12. R. Fandos, A. M. Zoubir and K. Siantidis, IEEE Trans. Geosci. Remote Sensing **52** (5), 2413-2426 (2014).