Multi-Aspect Acoustic Classification of Buried Objects

Marc Robinson, Mahmood R. Azimi-Sadjadi

Department of Electrical and Computer Engineering
Colorado State University
Fort Collins, CO 80523
azimi@engr.colostate.edu

Daniel D. Sternlicht and David Lemonds

ORINCON Defense
4770 Eastgate Mall
San Diego, CA 92121

Abstract- **In 2001, broadband echoes from objects buried 1-3 feet in a sand bottom were collected with the prototype of a Buried Object Scanning Sonar (BOSS), currently being developed by Florida Atlantic University and EdgeTech Inc. The object field consisted of one mine-like object as well as numerous non-targets, including scuba tanks, a concrete block, an aluminum sphere and an ordnance shell. The BOSS system was suspended in the water from a cable and translated across the object field. The task of detecting and classifying target objects is very challenging using only single-aspect acoustic returns. Features were extracted from these acoustic returns and then applied to various back-propagation neural network (BPNN) classifiers. A non-linear decision-level fusion was then applied to perform multi-aspect classification. The results are presented in terms of the receiver operating characteristics (ROC) curves and confusion matrices for the data collected by the BOSS system.**

I. INTRODUCTION

Acoustic detection and classification of buried sea mines presents a very challenging problem. Low acoustic frequencies are necessary for penetrating the top layer of ocean sediment (roughly < 30 kHz). At these frequencies, narrowband sonars have great difficulty in performing detection due to surface and bottom reverberation as well as spatial resolution limitations. The use of broadband transmission allows sonars to overcome this shortcoming.

There are many approaches for detection and classification of objects on or under the sea floor. One broadband acoustic approach to detection and classification of mines is outlined in [1]. With this approach high-frequency transmissions are used to delineate spatial characteristics of mine-size targets, by creating a pixel map of the sea floor. These characteristics include target size and shape, scattering strength relative to background, and the geometry of acoustic shadows. These characteristics are then used to classify the target objects. Another approach is to measure the acoustic echo return off the objects for a single sonar ping and investigate the spectral and temporal features of the return. Classifiers can then be trained using some of the spectral or temporal features to discriminate between targets and non-targets [2, 3]. Each approach has its own strengths as well as shortcomings. For example, classification from images requires a swath mapping sonar, but gives a wide area of coverage and a target's spatial characteristics (shape and size). Classification from only single aspect sonar returns, on the other hand, uses a very simple single beam sonar and requires far fewer aspects, but creates a very difficult signal processing problem because of the varying reverberation in each sonar ping. The performance can be enhanced by fusing together a small number of acoustic echoes (typically 3-5). This fusion can be performed in many ways, including feature-level fusion [4, 5], decision-level fusion [2] or a combination of both feature and decision-level fusion [6]. This paper investigates the performance of a multi-aspect fusion system that uses the acoustic echo returns.

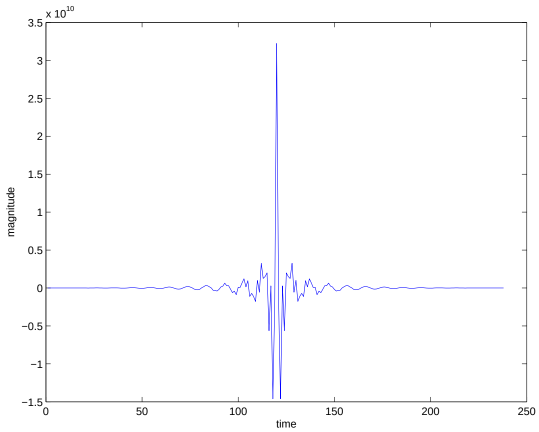
The data for this study was collected with the Buried Object Scanning Sonar (BOSS) developed by Precision Signal (now EdgeTech), Florida Atlantic University and Sea Engineering Inc [7]. The broadband transmission signal is a linear FM of duration 2 msec and bandwidth in the range of 5-23 kHz. Section II of this paper describes the BOSS system and the data collection process. The pre-processing and feature extraction steps are explained in Section III. The classification mechanism and fusion scheme are explained in Section IV, and the results on the BOSS data are presented. Conclusions are made in Section V.

II. BOSS SYSTEM AND DATA COLLECTION

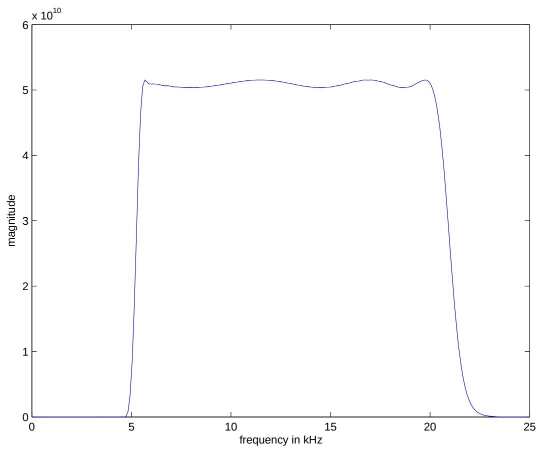
The BOSS [7] was designed to achieve imaging resolutions of 4 to 8 cm for objects buried 1 to 2 ft. below the seafloor. The system operates with a 2-msec, 5-23 kHz Linear FM transmit signal shown in Figure 1(a). The magnitude spectrum of this signal is shown in Figure 1(b), which demonstrates a flat behavior over the entire 5-23 kHz range.

The BOSS system consists of 6 transmitters, a 32 element receiver array arranged in 8 line arrays of 4 elements each, housed in a 300 pound tow vehicle 1.55 m (61 inches) long and 1.09 m (43 inches) wide. The line of 6 transmitters along the central axis is used to generate a projected sound that can be electronically steered. The system is capable of mapping the seafloor and sub-bottom at swath widths of two to four times the vehicle altitude (depending on the type of substrate). A controlled test range was designed and installed at Sea Engineering's facility at Makai Pier in Waimanalo, Oahu. The

0-933957-31-9



(a) Incident Signal



(b) Magnitude spectrum of the Incident Signal

Fig. 1. 5-23 kHz Incident Signal and Spectrum.

field consisted of objects buried between 1-3 feet below the seafloor, at various orientations and offsets to a 170-feet suspended cable stretched between two pilings. The water depth varies from 13 feet at one end of the field to 8 feet at the other end. Data collection is performed after targets were buried in the fine sand substrate for a period of almost two months. To collect sonar data, a hydraulically powered capstan translated the vehicle along the wire cable from one end of the test field to the other, during which pitch and roll were measured for each ping. The tow speed was approximately 0.7 knots and the ping repetition rate was 20/sec. Table 1 lists the objects buried in the seafloor, their composition and geometrical characteristics as well as the burial depth.

The acoustic data was sampled at 50 kHz. Multiple runs were taken in each direction along the field of targets. This was done to allow training in one run and testing in another, to examine the robustness of the features and the trained classifiers. For this study, two runs are used, both backward runs (deep to shallow end) with the electronically-steered beam angle of $0°$. The BOSS system takes swath measurements perpendicular to the direction of motion. Table 2 shows the number of aspects that were obtained for each object in each of the two runs. Clearly, the number of aspects depends on the length of the object as well as its orientation.

Table 1

.

BURIED OBJECTS IN THE BOSS DATA SET

Object Number	Description	Burial Depth (ft)
1	Concrete block 1.5 x 1.5 ft.	-1
4	Aluminum sphere, 1.5 ft diameter, filled with sand	-1
7,8	Steel scuba tank filled with concrete, 2 ft. x 7 in.	-1
9	Mine-shaped cylinder, 1.5 ft by 6.5 ft	-1.5
10	155mm ordnance, 0.5 ft x 3 ft	-1

Table 2

.

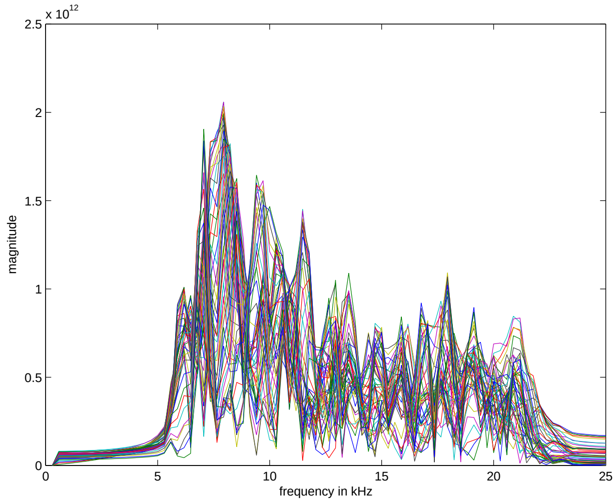
NUMBER OF ASPECTS AVAILABLE FOR EACH OBJECT IN EACH RUN

Target	Backward run 1	Backward run 2
1	51	51
4	51	51
7	61	61
8	61	61
9	131	131
10	71	71

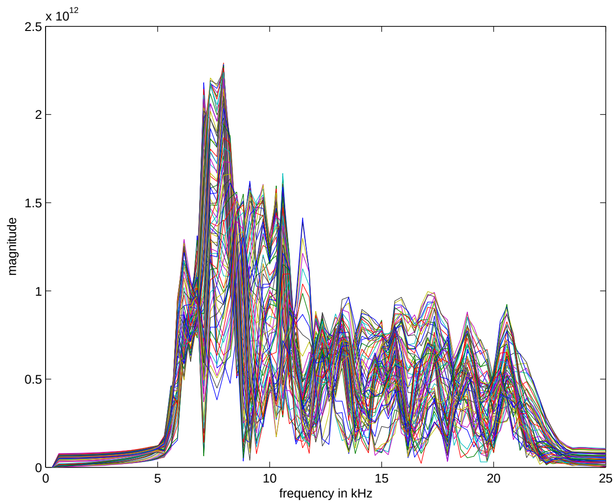
III. PRE-PROCESSING AND FEATURE EXTRACTION

The first procedure that must be performed is detection. This separates potential targets from background reverberation and natural structures within the sea floor (rocks, coral, etc.). Potential targets are found by using the backscatter image energy-clustering algorithm [7]. This algorithm integrates the energy reflected by an object over several transmissions and compares with a threshold. The data received had already undergone detection, so all aspects contain some type of object. The BOSS system is a swath mapping sonar and the 32 element receive array must be beamformed to create a single acoustic return. To maximize the SNR of the acoustic return, a focusing utility employs the spatial coordinates found by the detection process to extract the desired time-series. This utility is based on target locations, delay, transmit angle, vessel pitch and roll. Note that in this process the matched filtered returns from each receiver are focused. Additionally, the returns off the sea floor bottom must be removed, which is currently done manually.

All of the above steps were done by ORINCON [3]. Most of the backscattered energy associated with these targets occurs in the frequency band 5-15 kHz as can be seen in Figures 2(a) and (b), which show the magnitude spectra of the focused returns for all the aspects from Target 4 (aluminum sphere) for both backward runs. The additional attenuation at higher frequencies is caused by the sand substrate which causes a much larger attenuation at 23 kHz than it does for the lower frequencies (5-15 kHz). Figures 3(a) and (b) show the magnitude spectra of the focused returns from Target 9 (Mine-like object). In



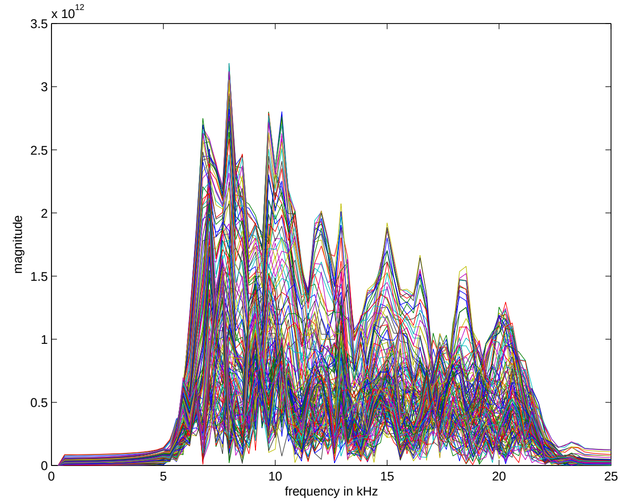
(a) First Backward Run



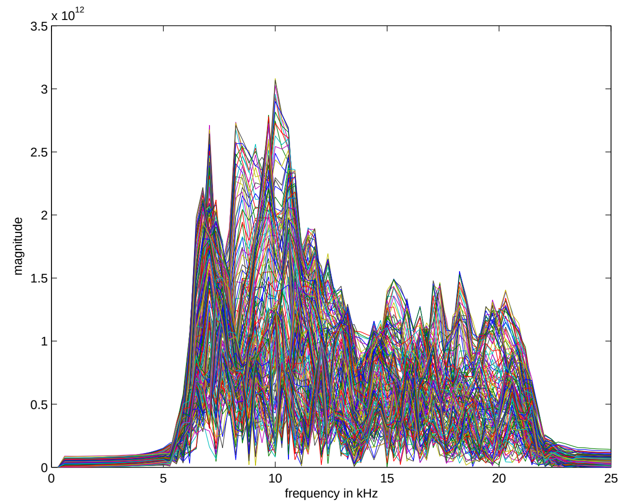
(b) Second Backward Run

Fig. 2. Magnitude Spectra of returns from Target 4 (Aluminum Sphere).

these two plots interesting behavior can be noted, namely at the lower frequencies. As can be observed, Target 4 exhibits minor low frequency changes from one run to another, while Target 9 undergoes substantial changes in these frequencies (5-15 kHz) from the first to the second backward run. In the first run for Target 9 (Figure 3(a)), there is a large peak at 7 and 8 kHz as well as two close peaks around 10-11 kHz. In the second run, the position and strength of these peaks changed. The peak at 7 kHz remains but becomes narrower, while the peak at 8 kHz has lost magnitude and is now combined with the other two peaks which were previously around 10-11 kHz. Additionally, the behavior at higher frequencies have changed as well. For frequencies in the range of 15-23 kHz, the first run has more noticeable peaks with larger magnitudes than the second run. From these observations one can conclude that the changes in



(a) First Backward Run



(b) Second Backward Run

Fig. 3. Magnitude Spectra of returns from Target 9 (Mine-like object).

the spectral features from one backward run to another are substantial for any classification system.

After the bottom return has been removed, 170 samples of the data remain for each aspect. This signal is then normalized by its Euclidean norm, so that the object strength is no longer used as a discrimination feature. Additionally, this compensates for the drastic changes in amplitude caused by different aspects on the objects.

It is not possible to extract statistics-based features from the time-series as the number of samples is too small. If wavelet packet (WP) [8] is used to decompose the signal into frequency sub-bands, the length of the signals in each sub-band goes down by a factor of 2 for each level of WP used. This makes it difficult to use time-series modeling methods, such as linear predictive coding (LPC) [9], to model the peaks in the frequency spectra in the sub-bands. Thus, there is a trade-off between frequency bin size and the length of the signal. In this study, a 3^{rd} order WP was used to decompose the acoustic returns into 8 sub-bands. A Symlet 4 wavelet was used due to its orthogonality and nearly linear phase characteristic. The length of the resulting time-series in each sub-band was 27 samples. Figure 4 shows the WP decomposition tree used and the resultant frequency sub-bands. The first sub-band was omitted as it was completely outside the bandwidth of the original transmit signal. Within each sub-band, the resulting time

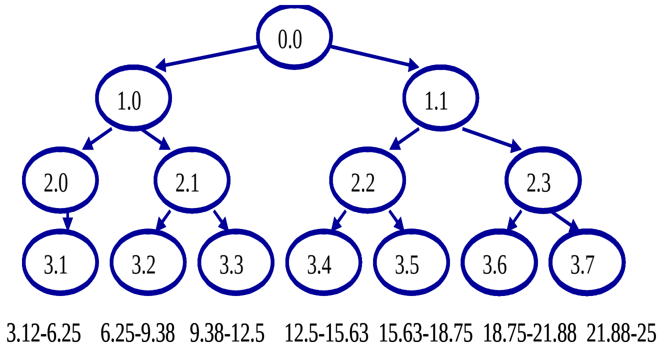


Fig. 4. Three-level Wavelet Packet Decomposition Tree.

series was zero-padded to twice its original length (54) and the FFT was taken. Only the first 27 points, owing to the periodicity of the FFT, were used as features for classification. Note that the magnitude of the FFT coefficients of the time series were used as features in order to eliminate any possible error or bias that could be caused by the initial onset of the objects return (which could vary depending on removal of the bottom return). This would also make the features to be independent of the burial depth. The entire system for pre-processing and feature extraction is shown in Figure 5.

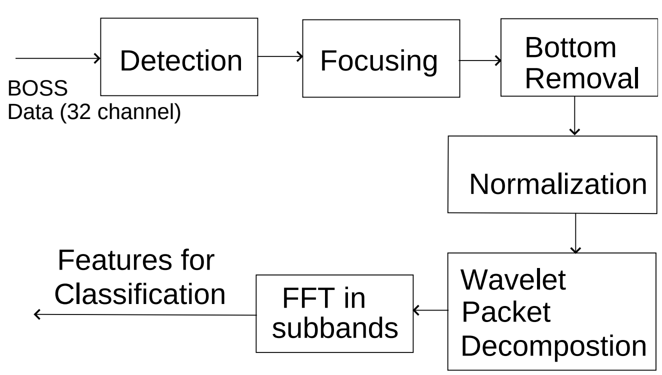


Fig. 5. Block Diagram of Pre-processing and Feature Extraction processes.

IV. CLASSIFICATION AND FUSION

The goal of this study is to see how well a classifier trained in one environmental condition (one backward run) performs in another environment (second backward run). This will be studied in each sub-band as well as after sub-band fusion and multi-aspect fusion. First, the data sets are split into training, validation and testing sets. Since this is a two-class problem, the training set should consist of equal number of target and non-target aspects. Thus, 120 samples of the target object (Target 9) and 30 samples of each non-target (Targets 1,4,8,10) within the first backward run were chosen randomly from all the available aspects in the run to form the training set. The validation set consisted of the even aspects for all the five objects in the second backward run and the testing set contained all the odd aspects of these objects on the second run. Target 7 (scuba tank) was used as an additional testing set, as no aspects from this target were included in the training set. This will enable us to understand the generalization ability of the trained classifier on this object.

Using the 27-dimensional spectral-based feature vector in each sub-band, several two-layer back-propagation neural networks were trained to perform two-class classification based only on the sub-band spectral features. The learning rate and momentum factor were .3 and .9, respectively. The error goal was 5 with a maximum number of epochs of 8000. Most of the networks met the error goal much fewer than the specified maximum number of epochs. Twenty different initializations were tried for each sub-band and the network that performed the best on the training and validation sets combined was chosen. The network structure was the same in each sub-band (27-54-2, i.e. 27 inputs, 54 hidden-layer neurons and 2 outputs, one for targets and one for non-targets), and multiple initializations were done in each sub-band for optimal training. Table 3 shows the correct classification rates on the training, validation and testing in each sub-band.

Table 3

.

CORRECT CLASSIFICATION RATES IN EACH SUB-BAND

Sub-band	Training	Validation	Testing
3.1	100%	73.8%	74.5%
3.2	100%	62.7%	61.6%
3.3	99.5%	78.3%	76.2%
3.4	85.0%	58.8%	55.6%
3.5	96.6%	60.0%	56.7%
3.6	97.0%	71.1%	70.8%
3.7	99.1%	69.4%	67.0%

As can be seen from Table 3, the training was good in all sub-bands except sub-band 3.4. The more interesting development is the extreme drop in the classification performance from one backward run to the other. For example, sub-band 3.2 drops from 100% training to only 62% on the second run. This was investigated in depth on Target 9, as it performed the worst out of all of the objects in the second run. The Fisher distance [10] was computed in each sub-band to find those features that are the most discriminatory between classes in the second run. The scatter plots of the top two discriminatory features are shown for both the training and the validation sets in Figures 6 and 7, respectively. These features correspond to the 11^{th} and 12^{th} magnitude spectrum components of the 3.3 sub-band FFT. As can be seen from these figures, the features of the target object undergo a change from the training

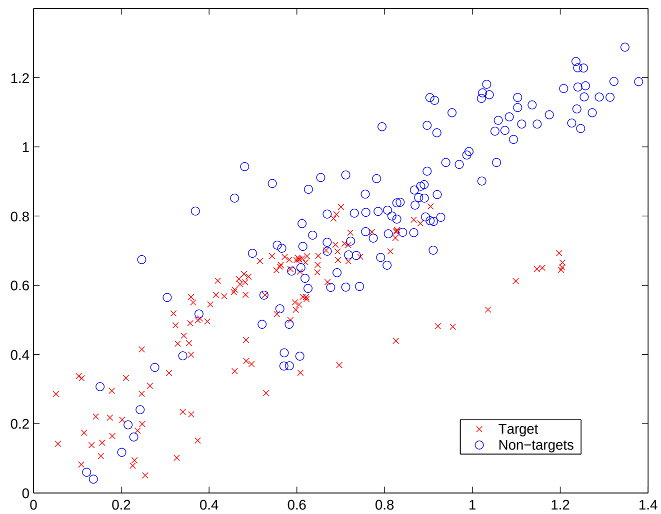


Fig. 6. Scatter plot of training set features.

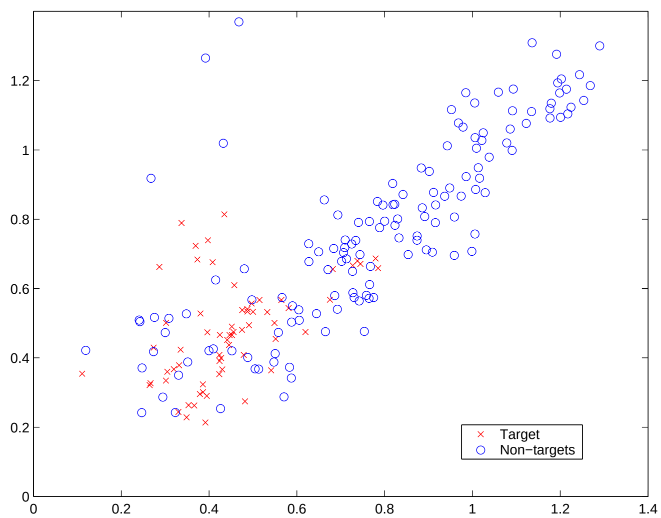


Fig. 7. Scatter plot of validation set features.

to the validation set. The features are much more compact in the validation set and have vacated many of the areas where there had been concentrations of training samples. These plots are good indications of why the performance suffers so dramatically from one run to another. The testing set looks nearly identical to the validation set, as they are taken from the same run.

A. Sub-band Fusion

After networks were trained in each sub-band, the outputs of these networks were then fused together using another BPNN. This fusion network was trained on the outputs of each sub-band network for the same training data. The structure of this two-layer neural network was 14-28-2. As expected the fusion network trained at 100%, as the training was at or near 100% in the individual sub-bands. Table 4 shows the results on each object after sub-band fusion. This classifier performs quite well

Table 4

CORRECT CLASSIFICATION RATE AFTER SUB-BAND FUSION

Object	Validation	Testing
1	88%	92.3%
4	88%	88.5%
8	96.7%	96.8%
9	58.5%	62.1%
10	88.6%	83.3%
overall	78.9%	80%
7	73.3%	77.4%

on the non-targets and struggles with the target. From Figures 6 and 7 the reason behind this observation is attributed to the movement of the Target 9 features from one run to the other. The receiver operating characteristics (ROC) curve in Figure 8 for the testing set exhibits probability of correct classification (P_{cc}) of 77.06% and a probability of false alarm (P_{fa}) of 22.96% at the knee-point of the curve. It is important to recall that the validation and testing sets do not have equal number of target and non-target aspects. As a matter of fact, the non-target aspects outnumber the target aspects 2:1. It is also interesting to look at the results of object 7 (second scuba tank), which was not included in the training set. It is the same size and shape as object 8, but has a different orientation, position and possibly a slightly different burial depth. The correct classification rate drops from 97% for object 8 to only 75% for object 7. This shows that the features indeed vary for the same object based upon the positioning of the object. Object 7 was parallel to the motion of the BOSS system and on the right side, while object 8 had an angle of $-135°$ to the direction of motion and was on the left side of the BOSS system. This would allow the BOSS system to have a bigger cross-section of object 8 within a single swath, and might explain the worse results on object 7.

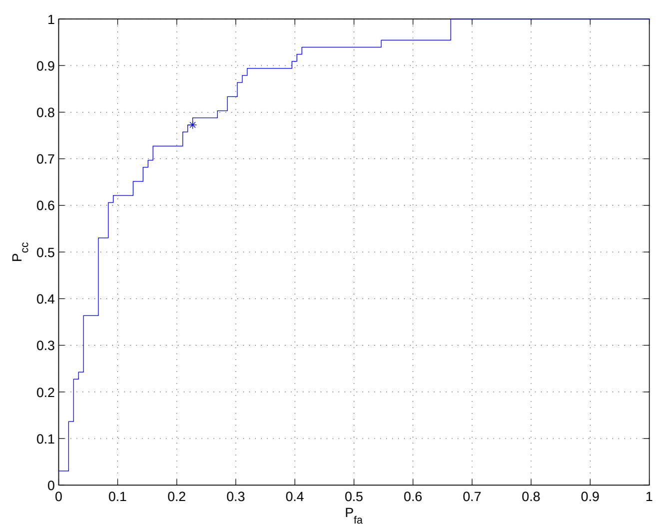


Fig. 8. ROC curve for the testing set.

B. MULTI-ASPECT FUSION

The next task is to take multiple aspects to allow for more information to be collected and to increase confidence in the decision. For this problem, consecutive aspects (pings) have an approximate distance of 0.2 inches between them, based upon the speed and the ping repetition of the BOSS system. In this case, 3 aspect fusion is accomplished after sub-band fusion has been performed for each aspect. A non-linear fusion scheme [2] was adopted here for its simplicity. The output of the single-aspect classifier for 3 aspects was taken and fused using a two-layer BPNN with the structure of 6-12-2. Training was done by taking 3 consecutive aspects from the single-aspect training set and forming a training set with 120 3-aspect combinations for both the target and non-target objects. Since the single-aspect training set was picked randomly from the available samples, consecutive samples correspond to an unknown amount of separation. Upon inspection of the training set, the largest separation for the target objects is 2 aspects (0.4 inches) and for non-targets the largest separation is 4 aspects (0.8 inches). The separation in the validation and testing sets is known, as they consist of every other aspect (even and odd respectively). The validation and testing sets were then formed into 3 aspect sets with about 0.8 (4 aspect separation) inches of separation between consecutive aspects.

After the BPNN was trained (at 100%), it was then validated and tested on the second run. Table 5 shows the correct classification rate after 3-aspect fusion for each object on the validation and testing sets. From Table 5 we can see that

Table 5

.

CORRECT CLASSIFICATION RATE AFTER MULTI-ASPECT FUSION

Object	Validation	Testing
1	84%	96.2%
4	96%	92.3%
8	100%	100%
9	80%	80.3%
10	85.7%	83.3%
overall	87.2%	88.1%

the multi-aspect fusion was able to account for many of the problems that the single-aspect classifier had particularly with the target object. The correct classification rate increased from about 60% to around 80% for Target 9. The improvement was much smaller for most of the non-targets. Nonetheless, the overall correct classification increased from 79% to 88%. Figure 9 show the ROC curves for the testing set after 3-aspect fusion. At the knee-point of this ROC curve we have $P_{cc} = 86.36\%$ and $P_{fa} = 13.64\%$.

To determine the optimum aspect separation interval, we varied the separation from 0.8 inches to 4.8 inches between consecutive aspects. It is important to note that the separation in the training set was also varied from approximately 0.8 to 4.8 inches. As can be seen from the plot of correct classification rate vs. aspect separation in Figure 10, the best rate is achieved at 1.6 inches for both the validation and testing sets

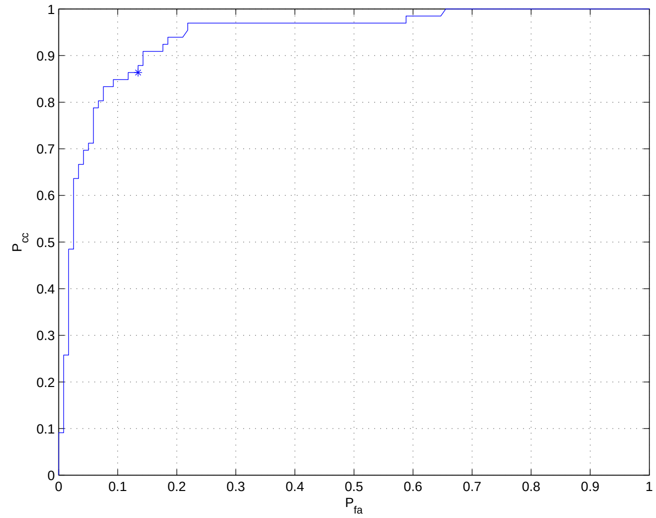


Fig. 9. ROC curve for the testing set after multi-aspect fusion.

where the correct classification rate is 90.6% and 88.6%, respectively. The jumps at 1.6 and 4 inches in the validation set are difficult to explain, and should be contributed to the characteristics of the data. This occurs when at a certain separation in the 3-aspect sets when fewer errors are grouped together than occurs at other separations. This multi-aspect fusion allows for a vast improvement over the single-aspect results with a very simple procedure.

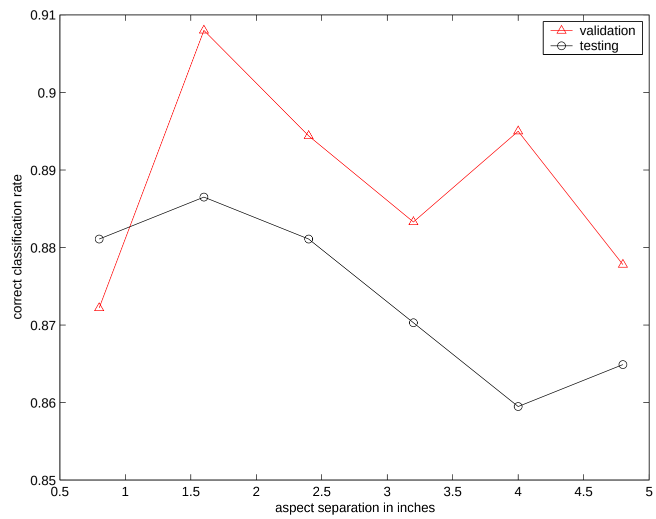


Fig. 10. Correct classification rate vs. aspect separation for multi-aspect fusion.

V. CONCLUSION

In this paper we have presented the results on the classification of buried objects from their acoustic returns. The objects were buried between 1 and 1.5 feet beneath the sea floor and remained there for over a month. The BOSS system was used to collect acoustic returns along the path that the objects were buried. These returns were then focused into a single acoustic time series based upon the likely position of the object under the sea floor (found within each sonar swath). The bottom return was removed from the matched filtered results. A 3^{rd} level WP was applied to decompose the time-series into sub-bands. An FFT was applied in each sub-band and the magnitude coefficients were used as features for classification. This is done in order to remove the dependency of features on the onset of the returns, especially after the removal of the bottom return. The classifier was a two-layer BPNN network trained based upon the training samples that were taken from one backward run of the BOSS system and validated and tested on another backward run.

The sub-band networks were successfully trained based upon the first run. Unfortunately, the features of the objects and in particular the target changed dramatically from one run to another. This could be caused by the pitch and roll of the vehicle, different environmental conditions (water temperature, turbidity), vehicle elevation from the bottom, or even a slightly different focusing point. This change causes the correct classification results to drop in each sub-band on both the validation and testing data sets (a difference of up to 30%). The sub-band results were then fused to create a final decision for each aspect. This fusion was done with a BPNN that gave a correct classification rate of about 79% on the second backward run. In this case, the single-aspect results were much better on the non-targets than they were with the target object. Indeed the P_{cc} for the target was very low at about 60%. To improve upon these results and increase confidence in the decision, 3 aspects were fused together. Again this fusion was performed by a separate BPNN. This approach was taken because of its simplicity. The multi-aspect fusion drastically improved the results on the target, increasing the correct classification rate from 60% to 80%. The non-targets also had a higher correct classification rate, but not to the same extent as the target. The overall classification rate improved by 9% on the second run, up to 88%. Considering the reverberation challenges imposed by burial in a sand substrate, these target classification results are very encouraging.

There is, however, much room for improvement. The first improvement would be to identify features that are more robust to environmental changes. It is interesting to note that if some aspects of both runs are used in the training set, that classification is excellent on both runs, even for a single aspect. Unfortunately we cannot expect to have training samples for all types of environments that we encounter. As a result, the challenge is to find features that are robust to changes in the environment, or to construct a classification system which adapts to new environmental conditions. Robust multi-aspect fusion schemes that may increase classification performance include: feature-level fusion [4, 5] and the combination of decision and feature-level fusion [6]. These schemes attempt to use the correlation between consecutive returns to improve the correct classification rate. Nonetheless, the simple sub-band fusion classification system presented in this paper performs very well on the buried target problem and the fusion of multiple aspects further improves upon the results.

ACKNOWLEDGEMENTS

This study was funded by the Office of Naval Research (ONR) under contracts #N00014-02-1-0006 and #N00014-02-C-0185. The data and technical support were provided by ORINCON Defense, San Diego, CA.

VI. REFERENCES

[1] G. J. Dobeck, J. C. Hyland and S LeDerick, "Automated Detection/Classification of Sea Mines in Sonar Imagery," in Detection and Remediation Technologies for Mines and Minelike Targets, Proceedings of SPIE 1997, Vol. 3079, pp. 90-108.

[2] M. R. Azimi-Sadjadi,De Yao, Qiang Huang and G. J. Dobeck, "Underwater Target Classification Using Wavelet Packets and Neural Networks," IEEE Trans. on Neural Networks, Vol. 11, No. 3, May 2000, pp. 784-794.

[3] D. D. Sternlicht, A. W. Thompson, D. W. Lemonds, R. D. Dikeman, M. T. Korporaal, "Image and signal classification for a buried object scanning sonar," Proceedings of Oceans '02 MTS/IEEE, Vol. 1, 2002 pp. 485-490.

[4] J. Salazar, M. Robinson and M.R. Azimi-Sadjadi, "Multi-Aspect Pattern Classification Using Predictive Networks and Error Mapping," Proceedings of the 2002 IEEE International Joint Conference on Neural Networks, May 2002, Vol. 2, pp. 1842-1847.

[5] J. Salazar, M. Robinson and M.R. Azimi-Sadjadi, "Multi-Aspect Discrimination of Underwater Mine-Like Objects using Hidden Markov Models," Proceedings of the Oceans 2002, MTS/IEEE Conference and Exhibits, Oct 2002, Vol. 1, pp. 46-53.

[6] M. R. Azimi-Sadjadi, D. Yao, A. A. Jamshidi and G. Dobeck, "Underwater Target Classification in Changing Environments using Adaptive Feature Mapping," IEEE Transactions on Neural Networks, 2002, Vol. 13, No. 5, pp. 1099-1111.

[7] S. G. Schock, A. Tellier, J. Wulf, J. Sash, M. Erickson, "Buried Object Scanning Sonar," IEEE Journal of Oceanic Engineering, 2001, Vol. 26, No. 4, pp. 677-689.

[8] S. G. Mallat *A Wavelet Tour of Signal Processing*, Academic Press, 1998.

[9] S. Haykin, *Adaptive Filter Theory, Third Edition*. Prentice Hall, 1996.

[10] R. O. Duda and P. E. Hart and D.G. Stork, *Pattern Classification*, A Wiley–Interscience Publication, 2001.