Featureless Classification for Active Sonar Systems

M. E. Soules and J. B. Broadwater
The Johns Hopkins University
Applied Physics Laboratory
Laurel, MD 20723 USA

***Abstract* – Active sonar systems depend on classification algorithms to identify target echoes and suppress false alarms. Historically, classifiers use a set of empirically derived features that exhibit some statistical separation between background clutter and target echoes. In an ideal scenario, these features would form a sufficient set of statistics capturing all of the information required to classify an echo return. Unfortunately, due to their empirically derived nature features are rarely provably sufficient. To overcome this drawback we present a featureless classifier. Instead of features, we use the raw data samples which form a trivial but provable set of sufficient statistics. To classify the raw data, we use the Adaptive Cosine Estimate algorithm which has a history of success with featureless classification in other applications. We will provide an in-depth look at our featureless algorithm that will include performance results on sea-test data from the Malta Plateau database.**

I. INTRODUCTION

Active sonar systems depend on classification algorithms to identify target echoes and suppress false alarms (i.e., clutter). Sources of clutter can include bathymetric features, the water surface, and various other objects in the environment such as wrecks. Additionally, radio frequency interference (RFI) on the communication link can introduce clutter. The role of the classifier is to separate the targets from the rest of these detected clutter events.

Historically, the development of classifiers for active sonar systems has been broken into four phases [1]. First, hundreds of features ranging from basic statistical approaches [2] to perception-based features [3] are created from model data and/or limited sea-test data. Second, a feature selection step is applied to reduce the number of features to reduce the effects of the curse of dimensionality [4]. Methods to perform this feature reduction have included analysis of individual feature performance, correlation between features, and use of metrics such as the multimodal overlap measure [1] or the Bhattacharyya distance [5]. Third, once the features are selected different classifier algorithms are implemented and their various parameters estimated from the training data. A good review of classifier algorithms can be found in [6][7]. Fourth, the various classifiers are compared using validation data – data that are intentionally held out of the training set. From this validation step, a final set of features and a classifier are chosen.

Because the traditional classifier only receives the feature values, it is dependent on the ability of features to extract the necessary information from the data stream. Ideally, a set of features should be a set of sufficient statistics. By definition, "a sufficient statistic for a parameter θ is a statistic that, in a certain sense, captures all the information about θ contained in the sample" [8]. Unfortunately, it is difficult to identify a set of features that is truly sufficient. While methods such as the multimodal overlap measure [1] or the Bhattacharyya distance [5] have shown promise, they cannot prove the set of features selected is sufficient. A better approach is to use the raw data samples directly in a classifier and bypass the standard feature extraction step. This raw data is a provable set of sufficient statistics (although not minimal) [8]. The difficulty with such an approach is identifying a classifier design that can handle such a large amount of data.

To identify the featureless classifier to use, we consider how other applications with similar issues handle classification. For example, face recognition algorithms have developed classifiers that operate directly on 512×512 pixel images and yet obtain 96% correct classification [9]. Similar results were also documented in radar [10] and hyperspectral remote sensing algorithms [11]. It is from this last application, we identified the Adaptive Cosine Estimate (ACE) algorithm for featureless classification [12]. Thus, instead of sending our classifier a vector of feature values estimated from the detection snippet, we are providing the entire data snippet to our featureless classifier ACE.

The rest of this paper is organized as follows. Section II describes the featureless classification algorithm including a derivation of ACE. Section III describes the sea-test data used for this research and the necessary steps taken to ensure independent training and test sets. Section IV demonstrates the capability of our featureless classifier to identify oil rigs in the Malta Plateau region [13]. Section V provides a summary of our work and future directions.

II. ALGORITHM DESCRIPTION

As with traditional sonar classifiers, the first step involves identifying events of interest in the data. A detector identifies and clusters together samples above a signal to noise ratio (SNR) threshold that represent echoes. For each of these echoes, a snippet is formed by extracting a fixed window of samples centered about the event including those that may be below the SNR threshold. Then, instead of computing various features from the data snippet, the featureless algorithm considers the entire snippet vector, $\mathbf{x}$, of beam time series data.

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

To reduce background noise, the snippet **x** is collected after standard signal processing techniques are applied to clean the signal. This additional processing helps the classifier focus on the signal characteristics that separate targets from clutter instead of focusing on environmentally sensitive noise.

After snippet extraction and signal processing is completed, the ACE algorithm is applied to **x**. ACE is based on the following statistical hypotheses:

$$\begin{aligned} H_0&: \mathbf{x} \sim N(0, \sigma_0^2 \mathbf{\Sigma}) \\ H_1&: \mathbf{x} \sim N(\mathbf{Sa}, \sigma_1^2 \mathbf{\Sigma}) \end{aligned} \tag{1}$$

where **x** is the $L \times 1$ snippet vector, $\mathbf{\Sigma}$ is an $L \times L$ covariance matrix estimated from the clutter training data, σ_0^2 and σ_1^2 are unknown scalars that account for SNR differences, **S** is a $L \times P$ target subspace obtained from the target training data, and **a** is a $P \times 1$ weight vector that models the current snippet as a linear combination of the target subspace **S**. An interesting point to note is that the H_1 hypothesis models a snippet as both a clutter return and a target return allowing for the fact that targets can be embedded in clutter.

To estimate the unknown parameters **a**, σ_0^2, and σ_1^2, we use maximum likelihood estimation techniques. The resulting unknown parameters are estimated as

$$\mathbf{a} = (\mathbf{S}^t \mathbf{\Sigma}^{-1} \mathbf{S})^{-1} \mathbf{S}^t \mathbf{\Sigma}^{-1} \mathbf{x}, \tag{2}$$

$$\sigma_0^2 = \frac{1}{L} \mathbf{x}^t \mathbf{\Sigma}^{-1} \mathbf{x}, \tag{3}$$

and

$$\sigma_1^2 = \frac{1}{L} (\mathbf{x} - \mathbf{Sa})^t \mathbf{\Sigma}^{-1} (\mathbf{x} - \mathbf{Sa}), \tag{4}$$

Substituting equations (2) through (4) into (1) and using a generalized likelihood ratio test, the ACE detection statistic is

$$d(\mathbf{x}) = \frac{\mathbf{x}^t \mathbf{\Sigma}^{-1} \mathbf{S} (\mathbf{S}^t \mathbf{\Sigma}^{-1} \mathbf{S})^{-1} \mathbf{S}^t \mathbf{\Sigma}^{-1} \mathbf{x}}{\mathbf{x}^t \mathbf{\Sigma}^{-1} \mathbf{x}} \tag{5}$$

which is scale-invariant and enjoys a theoretical constant false alarm (CFAR) property.

For the target subspace, **S**, we can use either all of the target echoes from the training data or we can use a subspace of the echoes obtained via singular value decomposition (SVD). An important note is that SVD is being used to reduce the number of target exemplars P in **S** – not the number of beam time-series samples L.

To select P, we use a simple automated method to determine the size of the subspace based on the directional energy captured by the singular values [14]. One uses the first P singular vectors corresponding to

$$\sum_{i=1}^{P} \left(\frac{\sigma_i}{T} \right) \geq K \tag{6}$$

where T is the sum of all singular values σ_i and K is the percentage of T desired (e.g., 99%).

In summary, the featureless classifier system performs the following steps:

1. Identify snippets using detection results.
2. Create snippet of data from signal processed data (e.g. post filtering and normalization).
3. Calculate the mean and covariance $\mathbf{\Sigma}$ from the clutter training data.
4. Subtract the clutter mean from all the data to match the zero-mean assumption in (1).
5. Calculate the target subspace **S** from the target training data using SVD and (6) to form the matrix.
6. Calculate classifier scores using (5) for each snippet in the test data (e.g. operational data).

III. Data Description

To test the ability of the proposed featureless classifier on active sonar signals, we use data from the Malta Plateau Clutter Track Database [13]. The database contains active acoustic tracks from the 1994 Shallow Water Active Classification (SWAC)–1 and the 1995 SWAC-2 sea tests. Across these two sea tests, twenty objects are tracked which include both man-made (e.g. wrecks and surface traffic) and geophysical (e.g. bathymetric features) entities.

Each track is comprised of multiple contact returns or snippets. For each snippet, two seconds and seven beams of data are extracted centered on the echo of interest. It is these snippets that will be used for both training and testing of the featureless classifier.

Of particular interest is the fact that the clutter population is affected by the limitations of only tracked objects being available in the database. We are restricted to those contacts with consistent enough returns over pings to establish tracks. Due to this limitation we are unable to demonstrate the potential of the featureless algorithm in identifying clutter snippets on individual pings that are not consistent enough to form a track (e.g. RFI). Only having clutter tracks presents a more challenging problem to the classifier since this type of clutter is generally more difficult to differentiate from targets.

Additionally, as noted in [13], there are some snippets within tracks that are falsely labeled as containing a detection when, in fact, there is minimal energy present. To be as unbiased as possible, we made no effort to remove these snippets.

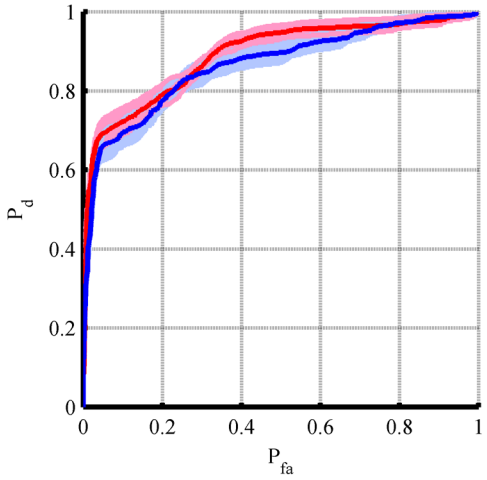

Fig 1. Featureless classification ROC curves for the single beam (blue line) and three beam (red line) data. Around each curve is the corresponding 95% confidence limits so the curves can be judged to be statistically similar or different.

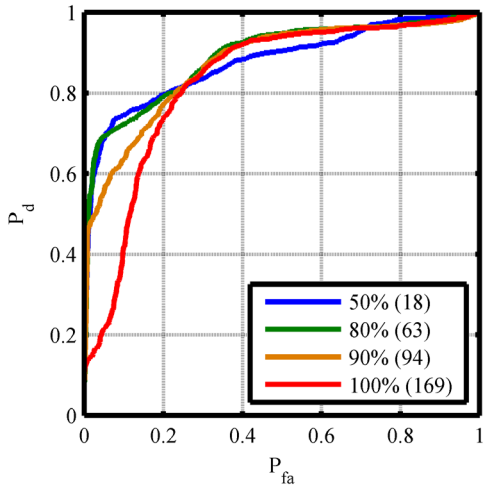

Fig 2. Featureless classification ROC curves for various target subspace dimensions. The percentage in the included legend represents the *K* value while the numbers in parentheses represent the corresponding *P* value in (3).

Acoustically, the database contains active returns from both low- and mid-frequency sonar systems. Therefore, numerous different types of coherent waveforms are available at a number of different frequencies. For most active sonar systems, tracks are constructed from wavetrains that contain multiple waveforms at different frequencies. For example, a typical wavetrain may contain a Hyperbolic Frequency Modulated (HFM) waveform followed by a Constant Wavelength (CW) waveform. The tracks in this database, however, are constructed from a single waveform whose frequency characteristics are constant over the track.

Based on the high number of tracks and snippets available across both SWAC-1 and SWAC-2, we chose to use the Linear Frequency Modulated (LFM) waveform tracks. Out of these data, we chose to use LFM waveforms with a center frequency of 600 Hz and a bandwidth of 400 Hz. Therefore the sample rate of the snippets is 667 Hz to allow for pass-band filter imperfections. Additionally, source levels for these tracks varied from 224 to 230 dB with an amplitude adjustment of 4 dB over the individual tracks.

From this data, we created our training and testing sets. For the training set, we use the first two tracks for each contact type. Two tracks for each oil rig are taken from the June 1995 SWAC-2 test to form the training target space. This provides us with 169 target returns for the training space and the remaining target tracks from both SWAC-1 and SWAC-2 tests provide 503 target echoes for our test data. Because not all clutter types have tracks in SWAC-2, we set aside the first two tracks for each clutter type regardless of the test. This provides 999 clutter detections for the training set and 3,175 detections for the test set.

An important fact to consider is that the training and testing sets were created using track level data – not snippet level data. Since tracks are typically comprised of multiple snippets that are at least partially correlated, we split the training data using the track level information. Had we used snippets, there would have most likely been snippets in the test and training data that came from the same track thus biasing the experimental results in the next section.

IV. Experimental Results

Three different experiments were used to demonstrate the usefulness of the featureless classifier. The first experiment was to compare performance of the classifier using the center beam of the snippet and the center beam plus its two neighboring beams. The idea with this experiment was 1) to demonstrate the usefulness of the featureless classifier and 2) to determine whether the featureless classifier could improve when given more information.

Figure 1 shows the Receiver Operating Characteristic (ROC) curves for the center beam versus the three beam data. The shaded areas depict the 95% confidence levels for each ROC curve. These confidence levels are included to identify when the ROC curves are statistically similar (curves overlap) and statistically different (curves are separated). The blue line provides the classifier performance for the center beam data and the red line provides the classifier performance for the three beam data. There is some gain in clutter reduction with the added beam information but overall the results are statistically similar. This gain can be seen when correctly classifying 93% of the targets. When using just center beam returns (blue line) the classifier keeps 65% of the clutter compared to the three beam data (red line) where the classifier only keeps 40% of the clutter – a 25 percentage point reduction. Other statistically significant gains can be found when the false alarm percentage is below 0.03%. At these levels, the three beam classifier correctly classifies nearly

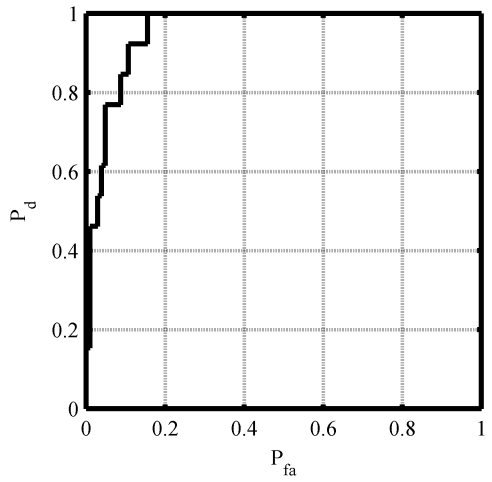

Fig 3. Featureless classification ROC curve for track level data.

double that found with the single beam classifier. In either case, the featureless classifier is performing well by maintaining high probability of detection (P_d) while suppressing a large percentage of the false alarms (P_{fa}).

The subspace dimension for the results in the first experiment used (3) with $K = 0.8$ resulting in a target subspace with $P = 49$. This matter of subspace dimension leads to the question of whether classifier performance is robust across different subspace dimensions estimated from equation (3). Therefore, the second experiment compares the different three beam ROC curves generated when K = 0.5, 0.8, 0.9, and 1.0.

Figure 2 provides the performance results for this second experiment. Of interest is the performance when K = 0.5 and 0.8. These curves provide the best performance from the four tested. As K gets larger, the performance decreases. This is most likely due to the target subspace becoming too large thus allowing in more clutter when applying a threshold for similar values of Pd. Another reason is due to the fact we did not remove snippets that had little or no acoustic energy due to label error. These snippets occur both in the target and clutter populations. When including all (or nearly all) of the target training snippets, these "null" snippets are also included which cause the performance to drop.

For the previous two experiments, we only showed performance at the snippet level. Since this data comes from track level information, it would also be interesting to see a ROC curve for tracks. To generate this ROC curve, we used the maximum classifier score over all snippets within a track as the track classification score. Those snippet classification scores were calculated using the 3 beam data with K set to 0.8. This method reduced our test set to 13 target tracks and 103 clutter tracks.

Figure 3 provides the track level ROC curve. As expected, this ROC curve is much better since it uses the best classifier scores from each track. This in effect removes any scores due to outliers or "null" snippets. The result is classifier

TABLE I
FALSE ALARM TRACKS FOR CLUTTER TYPE AT FIXED PD
(NUMBER OF FALSE TRACKS)/(TOTAL NUMBER OF TRACKS)

Clutter Type (Database Contact Number)	Pd for 13 Target Tracks		
	50%	80%	100%
Wreck (4)	2/5	3/5	3/5
Wreck (5)	0/7	0/7	0/7
System Transient (6)	1/25	2/25	4/25
Sediment Disturbance (7)	0/16	1/16	1/16
Wreck (8)	0/11	0/11	4/11
Wreck (9)	0/3	0/3	1/3
Rocky Outlier (10)	0/1	0/1	0/1
Possible Wreck (12)	0/6	1/6	1/6
Unknown (13)	0/1	0/1	0/1
Sediment Disturbance (19)	0/1	0/1	0/1
Random (20)	0/2	0/2	0/2
Ridge (21)	0/25	2/25	2/25
Total	3/103	9/103	16/103

performance where 100% of the target tracks can be maintained while rejecting 84% of the clutter tracks. This leaves only 16 clutter tracks in the database after featureless classification.

To identify which tracks remain, Table I lists the different clutter objects and how many false alarm tracks occur per object after classification. Wrecks make up over half of the clutter tracks with the rest coming from system transients and geophysical features. Since wrecks are submerged ships, the clutter from these objects is to be expected since they are similar to the oil rigs used as targets. If we remove these, only 7 clutter tracks remain that are significantly different from a submerged target track. Thus, featureless classification has performed well enough to reduce the population of true clutter below that of actual target tracks.

V. SUMMARY

We proposed a featureless classification system for active sonar systems. The algorithm is based on the well established ACE algorithm and standard SVD techniques. Experimental results demonstrate the ability of the featureless classifier to successfully separate targets from clutter even when the clutter makes meaningful tracks. By using SVD and (3), the target subspace is chosen to improve performance by eliminating "null" snippets and identifying the most robust target subspace to use.

While this paper documents the successful application of featureless classification to sonar, more work remains. First, the target subspace dimension P may require a more nuanced estimation technique such as Minimum Description Length [15]. Second, we need to apply the featureless classifier to data before tracks are created to measure its true performance. The idea behind a single-ping featureless classifier is to reduce the clutter to a level that clutter tracks are rare and target tracks can be created with less latency.

ACKNOWLEDGMENT

We first would like to thank Dr. Keith Davidson of ONR321 for sponsoring this research.

We would also like to acknowledge that the data used in this analysis were supplied by the Naval Undersea Warfare Center Division Newport, as acquired jointly with the NATO Undersea Research Centre, under the auspices of the Office of Naval Research, Code 321 Undersea Signal Processing.

REFERENCES

[1] D. H. Kil and F. B. Shin, *Pattern Recognition and Prediction with Applications to Signal Characterization*. Woodbury, NY: Springer, 1996.

[2] F. B. Shin, D. H. Kil, and R. F. Wayland, "Active Impulsive Echo Discrimination in Shallow Water by Mapping Target Physics-Derived Features to Classifiers," *IEEE Journal of Oceanic Engineering*, vol. 22, pp. 66-80, 1997.

[3] O. Ghitza, "Auditory Models and Human Performance in Tasks Related to Speech Coding and Speech Recognition," *IEEE Transactions Speech Audio Proc.*, vol. 2, no. pt. II, pp. 115-132, 1994.

[4] R. Bellman, *Adaptive Control Processes: A Guided Tour*. Princeton, NJ: Princeton University Press, 1961.

[5] A. Bhattacharyya, "On a Measure of Divergence Between Two Statistical Populations Defined by Probability Distributions," *Bull. Calcutta Math. Soc.*, vol. 35, pp. 99-109, 1943.

[6] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: John Wiley & Sons, Inc., 2001.

[7] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York: Academic, 1972.

[8] G. Casella and R. Berger, *Statistical Inference*. Belmont, CA: Duxbury Press, 1990.

[9] M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, vol. 3, pp. 71-86, 1991.

[10] L. M. Novak and G. J. Owirka, "Radar Target Identification Using an Eigen-Image Approach," in *Record of the 1994 IEEE National Radar Conference*, 1994, pp. 129-131.

[11] D. Manolakis and G. Shaw, "Signal Processing for Hyperspectral Image Exploitation," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 12-16, Jan. 2002.

[12] S. Kraut, L. Scharf, and R. Butler, "The Adaptive Coherence Estimator: A Uniformly Most-Powerful-Invariant Adaptive Detection Statistic," *IEEE Transactions of Signal Processing*, vol. 53, no. 2, pp. 427-438, Feb. 2005.

[13] W. V. Peterson and W. J. Comeau, "User Manual for Malta Plateau Clutter Track Database," Naval Undersea Warfare Center Division, 2007.

[14] L. N. Trefethen and D. Bau III, *Numerical Linear Algebra*. Philadelphia, USA: Society for Industrial and Applied Mathematics, 1997.

[15] J. Rissanen, "A Universal Prior for Integers and Estimation by Minimum Description Length," *Annals of Statistics*, vol. 11, pp. 416-431, 1983.