

CLASSIFICATION OF OBJECTS IN SYNTHETIC APERTURE SONAR IMAGES

Bradley Marchand, Naoki Saito, Hong Xiao

Department of Mathematics, University of California, Davis

ABSTRACT

This paper discusses an approach for the classification of objects in Synthetic Aperture Sonar (SAS) images and its benefit over other approaches. Our approach fully utilizes raw sonar waveforms scattered from objects. To do so, we first locate objects of interest in an image obtained by SAS processing. Then we extract the portions of the raw sonar waveforms responsible for forming those imaged objects from the whole raw sonar data. We align/straighten these extracted waveforms for localized discriminant feature analysis from which we obtain local features used for classification. We demonstrate the usefulness of our approach using real experimental sonar data.

Index Terms— Local Discriminant Basis, Synthetic Aperture Sonar, Pattern Classification

1. INTRODUCTION

In a typical Synthetic Aperture Sonar (SAS) image many objects are present. Some are of interest and others are natural materials that are not necessarily of interest. In particular, we are interested in the classification of underwater mines. We would like to classify objects in the imaged sonar data into a “mine” class or a “non-mine” class. Within the mine class we are also interested in the possibility of finer classifications (e.g., types of mines such as manta, etc.).

Currently, there are systems that are capable of generating 2-D and 3-D images of the sea floor and detecting objects on or below the surface of the sea floor. Sonar imaging systems have seen great improvements with the placement of the hydrophones and varying of the transmission beam. Such systems include SAS system [1, 2] and Buried Object Scanning Sonar (BOSS) system [3, 4]. There are a wide variety of techniques that are being developed to actually detect objects of interest. Many techniques rely on the signal strength, shape information of the imaged objects, or the amount of linear dependence (coherence) between two sonar returns of different grazing angles for the classification of objects [3, 5]. However there are several problems with using such information for classification. First, a buried mine will have a weaker signal than a proud mine. Second, a rock or some other non-mine objects could have a signal that is just as strong as a mine or stronger. Also, if one tries to use the shape information, the

shape can change with the strength of the return signal, orientation, and depth of the mine. Furthermore, other non-mine objects can have similar shapes to mines. Also, while coherence used in [5] will likely classify objects into the classes such as “simple geometric surface” and “irregular surface”, it is not likely to distinguish fake mines from real mines or be able to classify mines into finer classes. Hence, while such information is useful in determining which objects might be of interest, it may not be helpful with the classification of the objects.

In the next section, we propose a new approach that does not rely solely on the shape information of imaged objects but on the raw waveforms scattered from such objects. Then in Section 3, we demonstrate the capabilities of our approach by applying it to two different SAS datasets. Finally, in Section 4, we summarize our results and discuss possible future research directions.

2. OUR APPROACH

Our approach fully utilizes the scattered wavefield (i.e., the raw sonar waveforms) from objects of interest by facilitating the direct analysis of the raw waveforms. Furthermore, since such waveforms may contain material information (e.g., acoustic impedance, etc.), there is a potential to do finer classification of the objects.

To do so, our procedure first allows the user to interactively select and isolate objects of interest from SAS images of underwater scenes. Such SAS images are formed from the raw sonar data using an imaging algorithm (e.g., the $\omega - k$, the Range-Doppler, etc.; see [1, 2, 4] for the details). For an object of linear shape, we select two points representing the ends of the object. The points selected form a line representing the object selected. We then isolate and individually extract, from the raw sonar data, the waveforms responsible for forming the imaged object. This is done by the construction of an *envelope* (Figure 1). If we treat every point along the line corresponding to the object as a point source whose spreading is described by the equation

$$\left(\frac{tc}{2}\right)^2 = x_0^2 + (y - y_0)^2, \quad (1)$$

the envelope attempts to bound all such point sources. Here y is the cross-range location of the transmitter/receiver, t is the

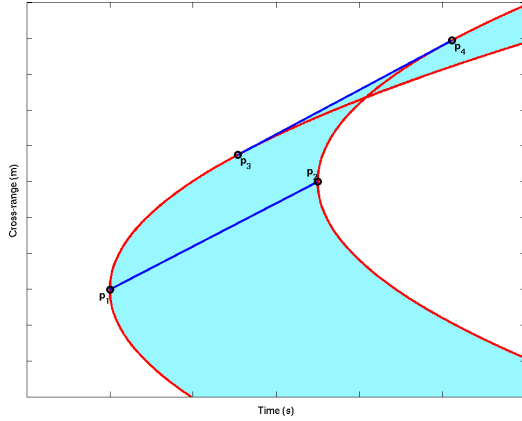


Figure 1: Envelope of oriented line. The points p_1 and p_2 are points selected by the user. The points p_3 and p_4 are used to form a line that will bound most of the point sources that are formed along the line connecting p_1 and p_2 . The shaded portion is the data that is extracted.

two-way travel time of the signal, (x_0, y_0) is the location of the point source, and c is the speed of sound through water. The envelope is formed from the hyperbolas corresponding to the point sources at the ends of the object and a line joining these hyperbolas. It is easy and straightforward to generalize this envelope computation for a line segment to that of a rectangle whose edges are parallel to the cross-range and range axes. By combining the envelopes corresponding to the left and right vertical edges, we have a very simple 2D envelope for the extraction of 2D objects.

Next, we align (or straighten) the extracted waveforms so that the time shifts of these waveforms are not used as a discriminating feature. We align the waveforms by applying a variation of the Range-Doppler algorithm [1, Chap. 4] which consists of the following steps: 1) Take the Fourier transform in the cross-range direction; 2) Apply the coordinate transformation

$$x(t, k_u) \triangleq \frac{c}{2}t \left[2 - \frac{1}{\sqrt{1 - \left(\frac{k_u}{2k_0}\right)^2}} \right]$$

$$k_y(t, k_u) \triangleq k_u$$

to the Fourier transformed data; 3) Take the inverse Fourier transform in the cross-range direction. Here k_u and k_0 are the cross-range and carrier frequency wavenumbers. The coordinate transform straightens the range migration and decouples the rows and columns of the range-Doppler matrix [8]. Further, the coordinate transformation reflects a scaling from a temporal to spatial ordinate [1, Chap. 4].

Once the extracted waveforms are aligned, we apply the

method called *Local Discriminant Basis* (LDB) [6, 7]. LDB automatically and computationally efficiently extracts local waveform features that are different among the signal classes of interest so that one can use these features to build a good classifier. The extracted local waveform features consist of localized amplitudes, phase, and frequency information at various time locations, which are often physically interpretable. These extracted features are then fed to a favorite classifier (e.g., linear discriminant analysis, neural networks, support vector machines, etc.).

3. RESULTS

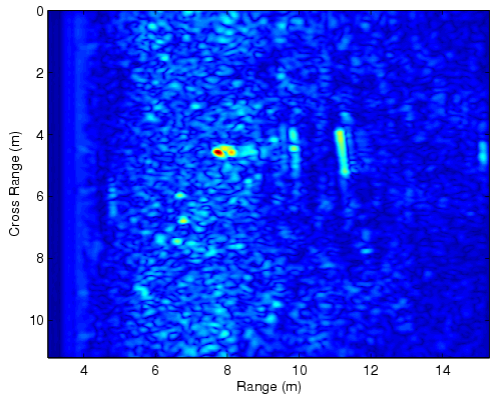
3.1. Dataset 1: Controlled experimental data in a test pond

As a first example, we consider data that was collected in a fresh water test pond at Naval Surface Warfare Center (NSWC), Panama City, FL. A sinusoidal pulse with a period of 0.2 ms and carrier frequency 20 kHz sampled at 500 kHz (i.e., $\Delta t = 2\mu s$) was used. There are 442 waveforms in total covering the 10.8 m of cross-range area. Each waveform has 8192 time samples, i.e., a time duration of 16.38 ms. There are two significant objects present in the image. The first is a 10 inch diameter steel sphere filled with air that is on the surface of the bottom of the pond at a range of approximately 7.6 m from the path of the transmitter/receiver and at cross-range about 4.5 m. The second, which is at a cross-range about 3.9 to 5.3 m, is a solid aluminum cylinder with a 30.5 cm diameter and 1.52 m length that is buried approximately 10 cm below the sand surface at a range of almost 10.3 m from the path of the transmitter/receiver. For more information see [9].

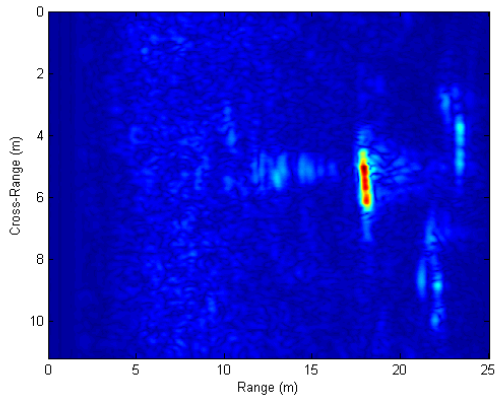
In Figure 2 (a) we show the imaged data. The size (or resolution) of each pixel is 0.15 cm along range by 2.54 cm along the cross-range. We extracted 81 waveforms corresponding to the cylinder and another 81 corresponding to the sphere and aligned them by the method described above. Figure 3 (a), (b) show the averaged aligned waveforms. Note that we extracted these waveforms from raw unprocessed sonar data (i.e., neither pulse compressed nor demodulated). The waveforms are clearly distinct.

In Figure 2 (b) we show another image where the cylinder was placed on the surface of the bottom of the pond. Let us call this dataset the “proud” dataset and the previous one the “buried” dataset for convenience. The proud data was recorded in a similar manner to the buried data, but on a different day. Moreover, a sphere was also present in this dataset but this sphere was different from the one in the buried dataset in two aspects: 1) the diameter is 14 inches; and 2) it contains silicone oil inside. Similarly to the buried cylinder dataset, we extracted and aligned 81 cylinder waveforms and 81 sphere waveforms.

Let $C1, S1, C2, S2$ be the set of the waveforms corresponding to the cylinder and the sphere in the buried dataset and those in the proud dataset, respectively. It is clear that



(a) Buried Cylinder



(b) Proud Cylinder

Figure 2: Buried and proud cylinders with spheres imaged with the $\omega - k$ algorithm.

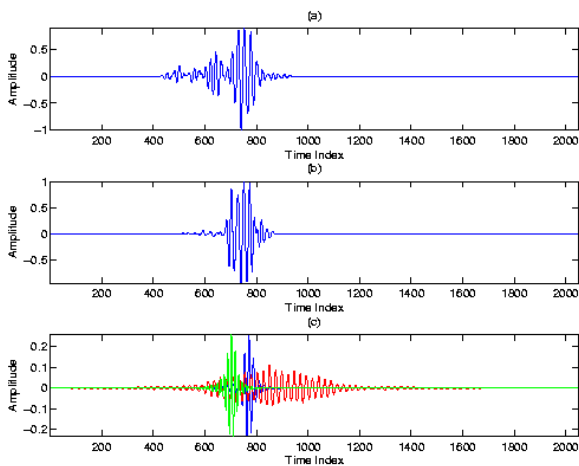


Figure 3: (a) Average cylinder waveform. (b) Average sphere waveform. (c) Top three LDB vectors for cylinder and sphere.

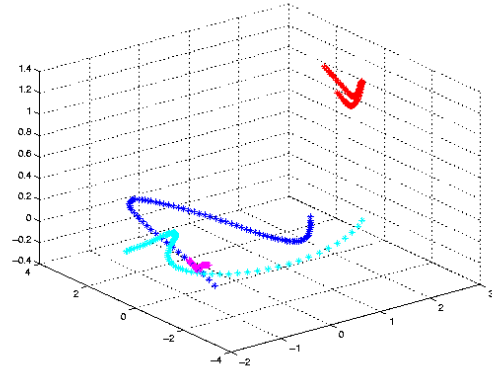


Figure 4: The top three LDB coordinates of the extracted waveforms. The LDB was computed on the $(C1, S1)$ dataset. Blue, red, cyan, and magenta points represent the waveforms belonging to $C1$, $S1$, $C2$, and $S2$, respectively.

we can form four possible training datasets, i.e., $(C1, S1)$, $(C1, S2)$, $(C2, S1)$, and $(C2, S2)$, on which we want to construct classifiers to discriminate the cylinder waveforms from the sphere waveforms and vice versa. The corresponding test datasets become: $(C2, S2)$, $(C2, S1)$, $(C1, S2)$, and $(C1, S1)$, on which we want to test the performance of the classifiers.

For each training dataset, we applied time-frequency LDB [6] algorithm with a symmetric discriminant measure (i.e., relative entropy) and 'Coiflet' 30-tap conjugate quadrature filter [10, Chap. 8] for developing a wavelet packet tree. We selected the three most discriminant LDB coordinates as the features to be used in classification. These three LDB vectors computed on the $(C1, S1)$ dataset are displayed in Figure 3 (c). We can easily see that specific localized time and frequency information of these LDB vectors indicates the distinguishing features of the cylinder and sphere waveforms whose averages are shown in Figure 3 (a), (b). For example, the 1st LDB vector (blue) checks the information around the time index 780 where the main energy of the both cylinder and the sphere waveforms are found. Around this time location, the energy and the phase of the sphere waveforms are quite uniform and coherent whereas those of the cylinder waveforms are slightly varying. The 2nd LDB vector (green) checks the information around the time index 700 where the difference between the cylinder and the sphere waveforms are quite clear from these figures. Figure 4 shows all the signals in these three LDB coordinates. It is clear that we can discriminate the cylinder signals (both $C1$ and $C2$) from the sphere signals in $S1$. However, those in $S2$ are not well separated from the rest.

Finally, for each classification experiment, we used Linear Discriminant Analysis (LDA) and Classification Tree (CT) as classifiers. Table 1 summarizes the misclassification rates of the test datasets.

| Test set | $C2$ | $S2$ | $C2$ | $S1$ | $C1$ | $S2$ | $C1$ | $S1$ |
|----------|------|------|------|------|------|------|------|------|
| LDA | 0.00 | 100 | 0.00 | 100 | 0.00 | 100 | 20.9 | 0.00 |
| CT | 1.23 | 100 | 0.00 | 100 | 0.00 | 100 | 0.00 | 100 |

Table 1: Misclassification rates (%) of the four experiments.

| Test set | $L2$ | $R2$ | $L2$ | $R1$ | $L1$ | $R2$ | $L1$ | $R1$ |
|----------|------|------|------|------|------|------|------|------|
| LDA | 59.5 | 52.1 | 39.7 | 38.8 | 83.5 | 11.6 | 60.3 | 28.1 |
| CT | 37.2 | 56.2 | 24.0 | 42.3 | 33.9 | 33.9 | 13.2 | 61.2 |

Table 2: Misclassification rates (%) of the four experiments for the field SAS data.

As we can see from Table 1, we could almost perfectly discriminate the cylinder waveforms regardless of whether buried or proud. On the other hand, we almost always failed to classify the sphere waveforms correctly. This is understandable since the acoustic characteristics of these two spheres are quite different (air vs silicone oil). The exception is the case of the test dataset ($C1, S1$) using the top three LDB vectors computed on the training dataset ($C2, S2$) with LDA as a classifier. In this case, the misclassification rate of the cylinder waveforms is 20.9% while that of the sphere waveforms is 0%, i.e., it was a perfect classification for the sphere, but not for the cylinder. This is because these two sphere waveforms happen to be on the same side of the LDA hyperplane in these top three LDB coordinates.

3.2. Dataset 2: Real field SAS data

For our second example, we use the SAS data that was collected at Buzzard’s Bay, Cape Cod. A ‘tone burst’ pulse with a carrier frequency 120 kHz sampled at 31 kHz was used. There are 900 waveforms in total covering the 18 m of cross-range area. Each waveform has 2016 samples, i.e., a time duration of 65.3 ms. The imaged data can be seen in Figure 5. Down the center of the image are several rocks. Among the rocks are two lobster traps. For this data set we do not have access to the raw unprocessed data. Therefore, we use the real part of the pulse compressed and demodulated data. We concentrate on two types of objects: lobster traps and rocks. We use the time-frequency version of LDB with the symmetric discriminant measure and the ‘Coiflet’ 12-tap conjugate quadrature filter for a wavelet packet tree. Waveforms for two lobster traps and rocks are extracted with an envelope for a rectangular object. Let $L1, L2, R1, R2$ be the sets of the waveforms corresponding to the two lobster traps and two rocks, respectively. For this field SAS dataset, we use the five most discriminant LDB coordinates for our classification experiments. The five LDB vectors computed on the ($L1, R1$) dataset are shown in Figure 6. The testing is done in the same way as our first example, and the results are shown in Table 2.

As the results indicate, it is much more difficult to classify objects from the pulse compressed and demodulated data. Contrary to the pond dataset, each waveform scattered from

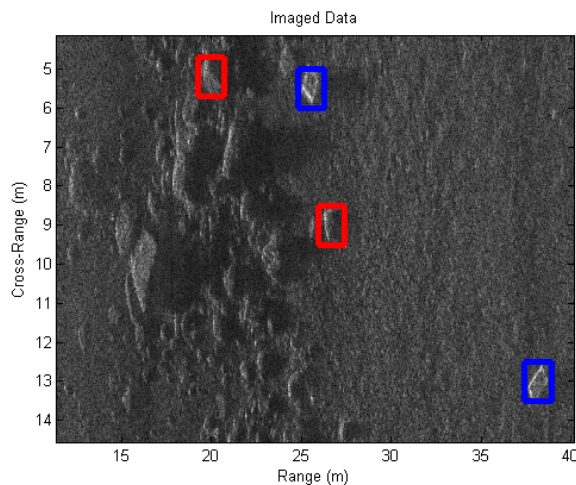


Figure 5: Imaged sonar data with objects labeled. The blue rectangles indicate the lobster traps while the red rectangles are rock regions.

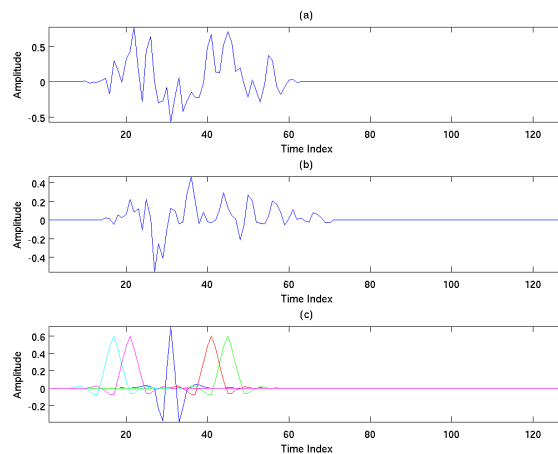


Figure 6: (a) Average lobster trap signal. (b) Average rock signal. (c) Top five LDB vectors for lobster trap and rock.

the same object varies considerably with respect to the others. This results in a complex distribution of the LDB coefficients, which can be seen in the large variation of results between LDA and CT. Linear hyperplanes are not able to separate lobster traps and rocks in these LDB coordinates. Further, the variation of results among training sets exhibits the need for a larger training set. The discriminant features selected vary greatly with the objects selected; thus, classification results are less stable between training sets.

Although classification results are expected to improve to some extent with a larger training set, a more fitted envelope (i.e., one for a rotated rectangle or a polygon), and a version of LDB accepting complex-valued waveforms, it is more important to take the whole scattered wavefield generated by an object into consideration when classifying. We will discuss this further in Section 4.

4. CONCLUSION

Our proposed approach could successfully discriminate the cylinder waveforms buried or proud from the sphere waveforms. The sphere waveforms could not be discriminated very well, which is understandable considering the difference in the sphere size and in particular the material inside of these spheres (air vs silicone oil): the acoustic characteristics between these two spheres are quite different. This in fact confirms our expectation that the raw sonar waveforms contain information about the material inside of targets. On the other hand, discrimination of the lobster traps from the natural rocks in the real field SAS data was quite difficult due to the considerable variation in time-frequency features in these waveforms.

In this paper, we use the local waveform features to classify and discriminate *each* waveform into two possible categories (i.e., cylinder vs sphere, or lobster trap vs rock). However, it is more natural to view the whole extracted waveforms corresponding to an object (i.e., the waveforms scattered from that object) as one signal (or more precisely an image). For example, we should view the cylinder dataset $C1$ as a single point in $\mathbb{R}^{2048 \times 81}$ rather than as 81 signals of length 2048. Therefore, we should investigate feature extraction, discrimination, and classification of the higher dimensional datasets. Our preliminary investigation along this direction can be found in our recent paper [11].

Finally, our current approach requires the user to locate and isolate objects of interest in SAS images interactively. We plan to investigate how to automate this object detection stage in the near future.

5. ACKNOWLEDGMENTS

We would like to thank Quyen Huynh, Dan Cook, Carrie Dowdy, Richard Holtzapple, and Joe Lopes at NSWC, Panama City, FL for their useful discussions and for providing us with

imaging software and real datasets to test our algorithm. This research was partially supported by ONR grants N00014-06-1-0615 and N00014-07-1-0166; and NSF grants DMS-0135345 and DMS-0410406.

6. REFERENCES

- [1] D. W. Hawkins, *Synthetic Aperture Imaging Algorithms: with application to wide bandwidth sonar*, Ph.D. thesis, Univ. of Canterbury, Christchurch, New Zealand, 1996.
- [2] H. J. Callow, *Signal Processing for Synthetic Aperture Sonar Image Enhancement*, Ph.D. thesis, Univ. of Canterbury, Christchurch, New Zealand, 2003.
- [3] S. G. Schock, A. Tellier, J. Wulf, J. Sara, and M. Erickson, "Buried object scanning sonar," *IEEE Journ. Oceanic Eng.*, vol. 26, no. 4, pp. 677–689, 2001.
- [4] S. G. Schock, J. Wulf, G. Quentin, and J. Sara, "Synthetic aperture processing of buried object scanning sonar data," in *Oceans 2005 MTS/IEEE Conference Proceedings*, 2005, vol. 3, pp. 2236–2241.
- [5] M. Yamada, J. Cartmill, and M. R. Azimi-Sadjadi, "Buried underwater target classification using the new BOSS and canonical correlation decomposition feature extraction," in *Oceans 2005 MTS/IEEE Conference Proceedings*, 2005, vol. 1, pp. 589–596.
- [6] N. Saito and R. R. Coifman, "Local discriminant bases and their applications," *J. Mathematical Imaging and Vision*, vol. 5, no. 4, pp. 337–358, 1995, Invited paper.
- [7] N. Saito, R. R. Coifman, F. B. Geshwind, and F. Warner, "Discriminant feature extraction using empirical probability density estimation and a local basis library," *Pattern Recognition*, vol. 35, no. 12, pp. 2841–2852, 2002.
- [8] R. Bamler, "A comparison of range-doppler and wavenumber domain SAR focussing algorithms," *IEEE Trans. Geoscience and Remote Sensing*, vol. 30, no. 4, pp. 706–713, 1992.
- [9] C. L. Nesbitt and J. L. Lopes, "Subcritical detection of an elongated target buried under a rippled interface," in *Oceans '04, MTS/IEEE Techno-Ocean '04*, 2004, vol. 4, pp. 1945–1952.
- [10] I. Daubechies, *Ten Lectures on Wavelets*, vol. 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*, SIAM, Philadelphia, PA, 1992.
- [11] L. Lieu and N. Saito, "Automated discrimination of shapes in high dimensions," in *Wavelets XII*, D. Van De Ville, V. K. Goyal, and M. Papadakis, Eds., 2007, vol. Proc. SPIE 6701, To appear.