

A New Synthetic Aperture Sonar Processing Method Using Coherence Analysis

Neil Wachowski, *Student Member, IEEE*, and Mahmood R. Azimi-Sadjadi, *Senior Member, IEEE*

Abstract—This paper introduces two methods that use a coherence analysis framework to generate synthetic aperture sonar (SAS)-like images that display acoustic color (AC) information useful for the classification of buried and/or proud underwater objects. The first method is applicable to sonar backscatter collected with multiple hydrophones and involves forming two channels using the data of two hydrophone subarrays at each frequency and over several pings. The second method is intended for applications where sonar backscatter is collected using a single hydrophone and involves forming two channels using the data of two synthetic subarrays at each frequency. In both cases, the resulting SAS-like AC images display information in a ping-frequency plane, and hence convey information that is useful for the detection, localization, and classification of underwater objects based on properties that are typically not conveyed by conventional SAS images. The single-hydrophone version of SAS-like AC processing has the added benefit of generating 3-D images that also display range information for enhanced localization capabilities. Furthermore, this coherence-based SAS-like method does not require the elaborate platform motion estimation and compensation used in conventional SAS. The effectiveness of these methods is demonstrated on two real sonar databases both by comparing generated SAS-like AC images to those generated using conventional methods and by applying a simple but effective classification framework directly to the AC features in the SAS-like images.

Index Terms—Buried object scanning sonar system, canonical correlations, synthetic aperture sonar (SAS), underwater target classification.

I. INTRODUCTION

SYNTHETIC APERTURE SONAR (SAS) processing [1], [2] is a popular method used to produce high-resolution images of the seafloor from which targets can be detected, localized, and possibly classified. SAS processing offers high azimuth resolution by coherently combining broadband sonar data from successive pings, thus making it an attractive system for underwater target detection and classification. In [3] and [4], SAS images are generated by coherently integrating the data from all the hydrophone elements on a linear array over a given number of pings to generate pixels that correspond to specific focal points on the seafloor. To account for near-field effects,

platform motion, and other sources of error, the samples from the different time series (recorded by each hydrophone element) that correspond to the same spatial location on the seafloor are found by calculating the differences in propagation time for each sensor element. Using a weighted sum of these samples, a 3-D matrix of pixels in the across-track, along-track, and depth dimensions is formed and projected onto a 2-D plane to generate a viewable 2-D image that represents the seafloor as viewed from one of three perspectives.

Many other ways of performing SAS processing besides the conventional delay-and-sum method have also been explored including those introduced in [5] and [6]. In [5], spatial resampling was employed so that narrowband adaptive beamforming algorithms could be applied to wideband active sonar returns collected with a small aperture uniform linear hydrophone array. Several narrowband adaptive beamforming algorithms, namely minimum variance distortionless response (MVDR), multiple signal classification (MUSIC), and spatial processing: optimized and constrained (SPOC) were then applied to real sonar data to generate images of the seafloor. Comparing the three adaptive beamformers, the SPOC algorithm provided the highest angular resolution, followed by the MUSIC algorithm, and finally the MVDR algorithm. Another study [6] modified several beamforming algorithms, namely the conventional delay-and-sum beamformer, the MVDR, the Fourier integral method (FIM), and the weighted FIM (WFIM), to make them suitable for processing data collected by a sparse platform mounted sonar. The results of applying each of these methods to real sonar data indicated the ability of the WFIM method to produce images with the highest SNR.

One common disadvantage of all the mentioned SAS processing methods is that there is often a large amount of ambiguity about the composition of the acoustical contacts depicted in the images they produce. In other words, an *object*, which may be defined as something insonified by the sonar that an operator is interested in characterizing, may appear similar to *clutter*, which the operator is generally not interested in, hence inhibiting the detection of the object. Traditional SAS images may be useful for localization and detection of underwater objects, and possibly classification based on an object's shape, but are typically not effective for classification based on other characteristics of an object, such as its composition. Thus, a conventional SAS image generated for an area of the seafloor that contains significant background clutter may have limited use for target classification, since objects in different classes often appear to be similar in the image. Moreover, many SAS processing methods require elaborate platform motion estimation and compensation techniques to account for the various sources of phase

Manuscript received June 04, 2009; revised February 18, 2011; accepted June 15, 2011. Date of publication August 30, 2011; date of current version October 21, 2011. This work was supported by the U.S. Office of Naval Research (ONR 321) under Contract N00014-10-1-0169.

Associate Editor: N. Chotiros.

The authors are with the Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523 USA (e-mail: nswachow@engr.colostate.edu; azimi@engr.colostate.edu).

Digital Object Identifier 10.1109/JOE.2011.2160471

errors. This requires employing rather expensive but accurate inertial navigation systems on the vehicle to monitor the platform path.

One approach for reducing composition ambiguity of the contacts captured by broadband sonar is to exploit the acoustical return off an object for a single sonar ping and investigate its temporal and spectral features. Acoustic color (AC) [7], [8] is a simple spectral-based feature extraction method that can capture additional information about the nature of acoustical contacts that is typically lost in SAS images. The method in [7] analyzes the difference between the spectral content in the transmit signal and relative absorption of each frequency in a return signal, i.e., the average absorption in each sonar return is removed. Using a collection of objects from different classes, the discriminating power of each frequency in these spectral density differences is then analyzed using the Fisher discriminant function [9]. Intensity levels for the red, green, and blue (RGB) channels in the color display are determined by integrating the energy in three corresponding frequency bands that are determined to be useful for object discrimination. In another study [8], a broadband AC display is generated from four images that represent the same area of the seafloor. One image is generated from the broadband data in a way that optimizes the detection of objects. The other three images are generated using low-frequency, mid-frequency, and high-frequency subbands of the broadband sonar data, that represent RGB component images, respectively. The image generated from broadband data is then used to determine the intensity of each pixel, while the three subband images are used to determine their hue and saturation values (converted from RGB components). The main problem with many existing AC processing methods is that they only use a single sonar ping (or beam) to extract the features, hence ignoring the ping-to-ping coherence information in generating the AC map.

To alleviate many of the shortcomings of conventional SAS and AC processing methods, this paper introduces a class of SAS-like AC processing methods. These methods do not require platform motion estimation and compensation, while at the same time generate images that convey information that is useful not only for underwater object detection and localization, but also for object classification. That is, SAS-like AC processing methods generate images that simultaneously have the properties of an AC image and a SAS image. This is accomplished by applying a coherence analysis framework [10], [11] to data channels that represent sonar data captured either by two different hydrophone subarrays in the case of multiple-hydrophone data, or two synthetic subarrays in the case of single-hydrophone data. The utility of these methods lies in the fact that the properties of different sources can be analyzed using canonical correlations [12] extracted from appropriately formed data channels [10], [11], and these correlation properties are often different in the presence of an object when compared with background clutter.

This paper is organized as follows. Section II introduces the SAS-like AC processing methods for multiple- and single-hydrophone data by outlining the data channeling process in each case followed by the image generation procedure. Section III describes the sonar data sets used to evaluate the proposed methods. Section IV presents the results of applying both

methods of SAS-like AC processing to the appropriate multiple-hydrophone and single-hydrophone data sets. Specifically, the visual properties of images generated using the proposed methods are first compared with those of conventional delay-and-sum and omega- k [13], [14] SAS images. This section also introduces and demonstrates a simple but effective method for detecting and classifying objects from SAS-like AC images generated from entire runs of the multiple-hydrophone sonar data. Finally, Section V concludes the studies carried out in this paper.

II. SAS-LIKE AC PROCESSING

SAS processing can be used to generate images of the seafloor that are typically useful for object detection and localization. Using a synthetic aperture improves the along-track resolution of target imagery with distance traveled and also allows the along-track length of the sonar hydrophone array to be significantly reduced, thereby reducing the hydrophone array drag and surface area and increasing the ease in which it can be deployed on unmanned underwater vehicles (UUVs). SAS processing is typically implemented by coherently integrating the data collected by one or more hydrophone elements on an array over a given number of pings to generate the pixels in an image. Coherent integration typically requires a preprocessing stage involving vehicle motion estimation and subsequent compensation achieved by employing inertial navigation systems on the vehicle to monitor the platform path. In contrast to conventional SAS processing, where coherency is achieved by estimating and accounting for the propagation time differences between the time series captured by each hydrophone element and ping [4] (or phase in the frequency domain), using canonical correlation analysis (CCA) [12] (see the Appendix) this coherency is achieved by linear transformation of the hydrophone data channels in the frequency domain. More specifically, the dominant canonical correlations of the mapped sonar data capture this coherency by correcting for all the propagation phases (delays) via the canonical coordinate mappings. Furthermore, since CCA is performed separately on different frequency bins of the sonar returns, as opposed to the entire time series, this SAS-like AC processing will produce images that reveal the spectral content at each ping, and hence, provide information useful for classifying objects that are proud, partially buried, or fully buried beneath the seafloor.

This section first discusses the channeling, vectoring, and averaging processes needed for subsequent two-channel CCA for multiple-hydrophone sonar data. The adaptation and modification of these processes to the case of single-hydrophone data is then outlined. In either case, the two-channel CCA is then applied and the resulting canonical correlations can be used to generate pixels for a SAS-like AC image. Note that both of these methods require that the impulse response of the object captured at each ping be computed for each hydrophone on the array. Extracting this impulse response can be accomplished by using the preprocessing method outlined in [11], which mainly consists of matched filtering each sonar return, eliminating the strong bottom return, windowing, transformation into the frequency domain, and removing the effects of the transmit signal via a pseudoinverse filter. Throughout this section, variables are

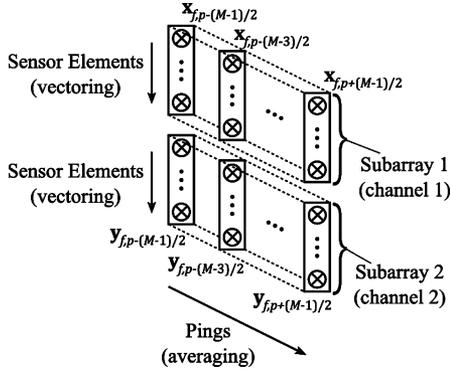


Fig. 1. Formation of CCA input channels for a single frequency f , when using multiple-hydrophone SAS-like AC processing.

marked with a $\tilde{\cdot}$ when they are associated with the single-hydrophone SAS-like AC processing.

A. Multiple-Hydrophone Data Channel Formation

In the multiple-hydrophone case, SAS-like AC processing is implemented by applying two-channel CCA to the data collected by two hydrophone subarrays observing the same area of the seafloor at a single frequency over a given number of pings. If the sonar that collected data consists of a single array of N hydrophones it can simply be manually divided into two subarrays, each containing $N/2$ hydrophone elements. The formation of the two input data channels for CCA is shown in Fig. 1. Each channel represents a separate hydrophone subarray and the elements of the vectors representing the channels are samples at the same frequency from the object's frequency response collected by the different hydrophones in the corresponding subarray. To compute the sample covariance matrices in the CCA (see the Appendix), averaging is done over a window of pings at the same frequency.

More explicitly, let us denote $\mathbf{x}_{fp} \in \mathbb{C}^{N/2}$ and $\mathbf{y}_{fp} \in \mathbb{C}^{N/2}$ as the complex vectors containing the zero-mean frequency response samples (all at the same frequency) obtained from the hydrophones on the first and second subarrays, respectively, at frequency bin f and ping p . To form the input channels for the two-channel CCA process M realizations of each subarray vector \mathbf{x}_{fp} and \mathbf{y}_{fp} are used, where each realization is formed using the data at a different ping in an odd-length window centered at ping p at frequency f (see Section II-C), the window of pings covers the range of $[p - (M - 1)/2, p + (M - 1)/2]$. This forms the input data matrices

$$X_{fp} = \begin{bmatrix} \mathbf{x}_{f,p-(M-1)/2} & \mathbf{x}_{f,p-(M-3)/2} & \cdots & \mathbf{x}_{f,p+(M-1)/2} \end{bmatrix} \in \mathbb{C}^{N/2 \times M} \quad (1)$$

$$Y_{fp} = \begin{bmatrix} \mathbf{y}_{f,p-(M-1)/2} & \mathbf{y}_{f,p-(M-3)/2} & \cdots & \mathbf{y}_{f,p+(M-1)/2} \end{bmatrix} \in \mathbb{C}^{N/2 \times M} \quad (2)$$

for subarrays 1 and 2, respectively. Note that here it is assumed that the frequency samples $\mathbf{x}_{f,p+i}$ and $\mathbf{y}_{f,p+j}$, $i, j \in [-(M - 1)/2, (M - 1)/2]$, obtained by the hydrophone

elements for each of the M pings, form independent realizations of the same random vectors. This also requires that the distance of each hydrophone to the seafloor does not change significantly from ping to ping, since this would cause large variations in the phase of the samples that averaging is performed over. Also note that the number of pings used for averaging M must be large enough relative to the subarray size $N/2$ to avoid generating sample-poor covariance matrices that may be deficient in capturing the canonical correlations and coordinates, i.e., we must ensure $N/2 \ll M$. This must be balanced with the fact that using too large of a value for M forms canonical coordinate mapping matrices that incorporate information from too large of an area of the seafloor, and hence, produces dominant canonical correlations that incorporate statistics of the data outside a localized area.

B. Single-Hydrophone Data Channel Formation

When data from only one hydrophone are available the CCA data channels cannot be formed using the above-mentioned approach. Therefore, in this case, a synthetic aperture is formed over the data from several pings, as it is commonly done in SAS processing [4]. However, this implies that averaging over ping data to form the necessary covariance matrices is no longer an option, and hence, we must find a new way to obtain different realizations of the same frequency to average over. This is done by segmenting the impulse response of the object (and bottom if they are overlapping) returns into L separate time segments (see Fig. 2), zero-padding each segment to the length of the original time series, and transforming each zero-padded segment into the frequency domain. The samples corresponding to the same frequency from these L segments are then used for averaging to compute sample covariance matrices needed in the CCA and to generate the canonical coordinate mapping matrices \tilde{F}_{fp} and \tilde{G}_{fp} [see (23) in the Appendix]. Each segment is then separately mapped using these canonical coordinate mapping matrices, hence preserving the correspondence between the samples and the range cells they represent. As will be shown later, the benefit of this mapping method is that the contribution of each range segment to the overall coherence of the data from the synthetic channels can be identified. This allows generation of SAS-like AC images that also provide localization information in the direction that the sonar was pointed; e.g., for side-looking sonar, across-track localization information can also be displayed.

As shown in Fig. 2, a single realization of the first data channel $\tilde{\mathbf{x}}_{fp}(l) \in \mathbb{C}^P$, where P is the number of pings used to form one synthetic subarray, contains the frequency samples obtained from the l th segment of the impulse response captured by the lone hydrophone at frequency bin f and pings $[p, p + P - 1]$. Similarly, $\tilde{\mathbf{y}}_{fp}(l) \in \mathbb{C}^P$ contains the frequency samples obtained from the l th segment of the impulse response captured by the same hydrophone at frequency bin f and pings $[p + J, p + J + P - 1]$, where J is the ping separation used as the displacement between the two synthetic subarrays. The input data matrices $\tilde{X}_{fp} \in \mathbb{C}^{P \times L}$ and $\tilde{Y}_{fp} \in \mathbb{C}^{P \times L}$ for the single-hydrophone SAS-like AC processing can then be formed

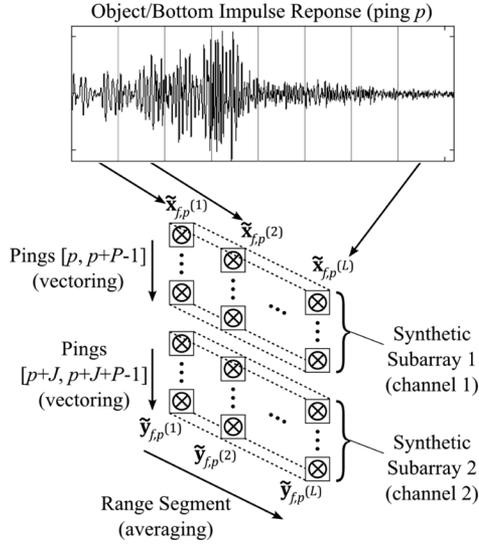


Fig. 2. Formation of CCA input channels for a single frequency f , when using single-hydrophone SAS-like AC processing.

using all L of the independent realizations of $\tilde{\mathbf{x}}_{fp}(l)$ and $\tilde{\mathbf{y}}_{fp}(l)$ obtained from all the segments of the impulse response, i.e.,

$$\tilde{\mathbf{X}}_{fp} = [\tilde{\mathbf{x}}_{fp}(1) \quad \tilde{\mathbf{x}}_{fp}(2) \quad \cdots \quad \tilde{\mathbf{x}}_{fp}(L)] \quad (3)$$

$$\tilde{\mathbf{Y}}_{fp} = [\tilde{\mathbf{y}}_{fp}(1) \quad \tilde{\mathbf{y}}_{fp}(2) \quad \cdots \quad \tilde{\mathbf{y}}_{fp}(L)]. \quad (4)$$

C. Generating a SAS-Like AC Image

Once the data channels have been formed in either the multiple-hydrophone or single-hydrophone case, two-channel CCA is used to compute the canonical coordinates and correlations of the frequency response at ping p and frequency f . This requires computing the sample autocovariance and cross-covariance matrices of the input data channels (shown below for the multiple-hydrophone case)

$$R_{\mathbf{xx}}^{(f,p)} = X_{fp} X_{fp}^H \quad (5)$$

$$R_{\mathbf{yy}}^{(f,p)} = Y_{fp} Y_{fp}^H \quad (6)$$

$$R_{\mathbf{xy}}^{(f,p)} = X_{fp} Y_{fp}^H \quad (7)$$

where H is the Hermitian operator. The sample coherence matrix is then computed using (23) in the Appendix as

$$C_{fp} = (R_{\mathbf{xx}}^{(f,p)})^{-1/2} R_{\mathbf{xy}}^{(f,p)} (R_{\mathbf{yy}}^{(f,p)})^{-H/2} \\ = F_{fp} K_{fp} G_{fp}^H \quad (8)$$

where the term on the right-hand side is its singular value decomposition (SVD) and $K_{fp} = \text{diag}[k_{fp}^{(1)} \quad k_{fp}^{(2)} \quad \cdots \quad k_{fp}^{(N/2)}]$ is the canonical correlation matrix. The orthogonal canonical coordinate mapping matrices F_{fp} and G_{fp} are used to map the input data channels X_{fp} and Y_{fp} to their respective canonical coordinates U_{fp} and V_{fp} , i.e.,

$$U_{fp} = F_{fp}^H (R_{\mathbf{xx}}^{(f,p)})^{-1/2} X_{fp} \quad (9)$$

$$V_{fp} = G_{fp}^H (R_{\mathbf{yy}}^{(f,p)})^{-1/2} Y_{fp}. \quad (10)$$

Let us denote $\mathbf{u}_{fp}^{(n)}$ and $\mathbf{v}_{fp}^{(n)}$ as the n th row of the canonical coordinate matrices U_{fp} and V_{fp} , respectively. Then, $\mathbf{u}_{fp}^{(n)}$ and $\mathbf{v}_{fp}^{(n)}$ contain samples of the n th canonical coordinates of the input channels X_{fp} and Y_{fp} , respectively. The first pair of canonical coordinates $\mathbf{u}_{fp}^{(1)}$ and $\mathbf{v}_{fp}^{(1)}$ has the highest correlation (or coherence) possible between the two sets of coordinates U_{fp} and V_{fp} . In the case of multiple-hydrophone data, this principal correlation can be estimated using

$$k_{fp}^{(1)} \approx \sum_{m=p-(M-1)/2}^{p+(M-1)/2} u_{fp}^{(1)}(m) v_{fp}^{(1)*}(m) \quad (11)$$

where

$$u_{fp}^{(1)}(m) = \mathbf{f}_{fp}^{(1)H} (R_{\mathbf{xx}}^{(f,p)})^{-1/2} \mathbf{x}_{fm} \quad (12)$$

$$v_{fp}^{(1)}(m) = \mathbf{g}_{fp}^{(1)H} (R_{\mathbf{yy}}^{(f,p)})^{-1/2} \mathbf{y}_{fm} \quad (13)$$

and $\mathbf{f}_{fp}^{(1)}$ and $\mathbf{g}_{fp}^{(1)}$ are the first column of F_{fp} and G_{fp} , respectively. To generate a pixel that represents ping p , only the contribution of coherence associated with a small, odd-length window of Q pings ($Q \ll M$) that are local to ping p will be used. Specifically, the pixel intensity for frequency f and ping p in the multiple-hydrophone SAS-like AC image $A(f, p)$ is found using

$$A(f, p) = \left| \frac{1}{Q} \sum_{m=p-(Q-1)/2}^{p+(Q-1)/2} u_{fp}^{(1)}(m) v_{fp}^{(1)*}(m) \right|. \quad (14)$$

An image is formed by generating pixel intensities using different samples of the frequency response, shifting the ping window, and repeating this process for all frequencies and pings. The window size Q should be small enough so that the coordinate samples used correspond to pings that are in close proximity to ping p , otherwise a blurred image may result due to incorporating pings that do not capture the same area of the seafloor.

In the case of single-hydrophone data, this principal correlation is estimated using

$$\tilde{k}_{fp}^{(1)} \approx \sum_{l=1}^L \tilde{u}_{fp}^{(1)}(l) \tilde{v}_{fp}^{(1)*}(l) \quad (15)$$

where

$$\tilde{u}_{fp}^{(1)}(l) = \tilde{\mathbf{f}}_{fp}^{(1)H} (\tilde{R}_{\mathbf{xx}}^{(f,p)})^{-1/2} \tilde{\mathbf{x}}_{fp}(l) \quad (16)$$

$$\tilde{v}_{fp}^{(1)}(l) = \tilde{\mathbf{g}}_{fp}^{(1)H} (\tilde{R}_{\mathbf{yy}}^{(f,p)})^{-1/2} \tilde{\mathbf{y}}_{fp}(l) \quad (17)$$

with all variables being defined similarly to their multiple-hydrophone counterparts. Since $\tilde{u}_{fp}^{(1)}(l) \tilde{v}_{fp}^{(1)*}(l)$ represents the contribution of coherence associated with the l th range cell segment to $\tilde{k}_{fp}^{(1)}$, it is used as the voxel intensity for frequency f , ping p ,

and range segment l in the 3-D SAS-like AC image $B(f, p, l)$, i.e.,

$$B(f, p, l) = \tilde{u}_{fp}^{(1)}(l) \tilde{v}_{fp}^{(1)*}(l). \quad (18)$$

An image is formed by generating pixel intensities using different samples of the frequency response from each range segment, shifting the ping window that forms the synthetic aperture, and repeating this process for all frequencies and pings. This SAS-like AC image offers classification information (f -plane) along with localization in two dimensions (p -plane and l -plane).

III. DATA DESCRIPTION

This section describes two data sets that are used to perform the experiments in this study. The buried object scanning sonar (BOSS) system and its associated data set are covered first. This data set is used for evaluating the performance of the multiple-hydrophone SAS-like AC processing including visual properties of the images it produces, localization capabilities, and explicit classification performance. The Ex10 pond system and its associated data set are then introduced. This particular data set is used for evaluating single-hydrophone SAS-like AC processing in the same manner as the multiple-hydrophone method evaluation, with the exception that explicit classification results are not generated for this method because the classifier is only applicable to 2-D images.

A. BOSS System and Collected Data Set

The wing BOSS was designed [4] to scan for buried underwater objects using two 1-m hydrophone arrays mounted on the wings of the Bluefin 12 UUV [15]. Each wing contains 20 hydrophones arranged in a uniform linear array, yielding a 40-hydrophone array used to collect the sonar returns. This system uses an omnidirectional projector that transmits a linear-frequency-modulated (LFM) signal over 3–19 kHz. The data set used in this study was collected off the coast of Panama City, FL in March 2007 using the wing BOSS system, and contains a mixture of target and nontarget objects. Fig. 3 shows the relative position of each object in the target field, while Table I gives the list of the objects, their properties, and burial conditions whether proud or buried. Note that, although there are objects with varying dimensions and compositions within the same class as well as objects in different classes with similar dimensions and compositions, there are often subsets of AC features (pixels) that are similar for objects in the same class, thus allowing for discrimination between objects in the target and nontarget classes. This property is exploited in the classification framework introduced in Section IV-B. During the data collection phase the Bluefin 12 UUV operated at a speed of 1.5 m/s and used a ping rate of 25 pings/s. The vehicle altitude was generally between 2.25 and 3.00 m, and the target field bottom was smooth and sandy. The collected sonar data were sampled at 40 043 Hz. For more information on the BOSS data set, see [11].

Sonar returns off objects in the data set were collected over a variety of runs through the target field. In particular, the Bluefin 12 UUV made five groups of runs that centered around objects T1, T3, T4, T5, or T10. These objects were chosen as the centers of groups mostly because their locations in the target field

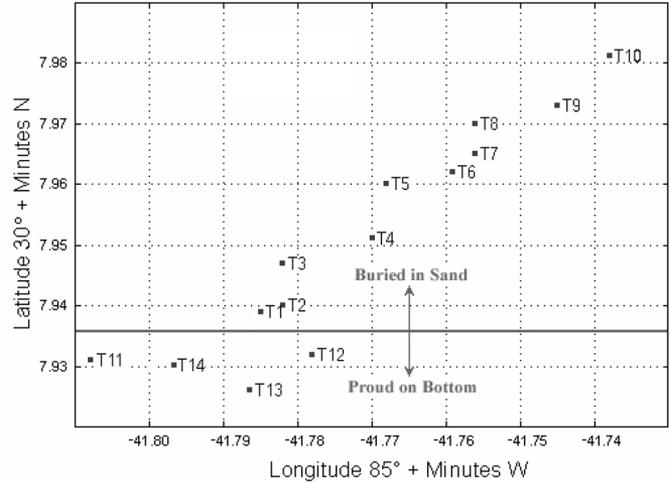


Fig. 3. BOSS target field layout.

TABLE I
DESCRIPTION FOR EACH OBJECT IN THE BOSS DATA SET.
OD = OUTSIDE DIAMETER

Label	Description	Status
T1	1.83 m Bomb-Shaped Marker (0.46 m OD)	Buried
T2	0.51 m Artillery Shell	Buried
T3	1.68 m Bomb-Shaped Target (0.28 m OD)	Buried
T4	1.83 m Iron Cylinder (0.46 m OD)	Buried
T5	1.52 m Iron Cylinder (0.18 m OD)	Buried
T6	0.36 m Stainless Steel Sphere	Buried
T7	0.89 m Artillery Shell	Buried
T8	0.36 m Stainless Steel Sphere	Buried
T9	0.61 m Iron Cylinder (0.15 m OD)	Buried
T10	1.83 m Bomb-Shaped Marker (0.46 m OD)	Buried
T11	2.44 m Concrete Pipe (0.46 m OD)	Proud
T12	1.83 m Concrete Pipe (0.46 m OD)	Proud
T13	1.83 m Concrete Pipe (0.46 m OD)	Proud
T14	1.83 m Concrete Pipe (0.46 m OD)	Proud

allow a single run over them to capture sonar returns from a large number of other surrounding objects. These groups consist of runs that were made in north–south, east–west, and north–west–southeast trajectories at varying distances from the center object. This method of data collection allowed pings from different runs to capture signatures from different aspects of each object. In addition, the Bluefin 12 UUV made diagonal runs in a northeast–southwest trajectory over the entire target field, capturing one to four objects per run. These diagonal runs not only obtained more data on all the objects covered by the other five groups, except T12 and T13, but were also able to capture sonar returns from objects not covered by the other five groups, namely those of T11, T14, T7, and T8. A run typically captures a large number of sonar pings that are not off an object. For instance, out of typically 500–1400 pings for a run, only 8–30 pings on average cover an object of interest, depending on the orientation of the object relative to the sonar platform. Each sonar ping was preprocessed using the procedure in [11] before the application of the SAS-like AC processing. As indicated before, this preprocessing entails ideally removing portions of each sonar return that do not represent insonified objects. Specifically, this involves applying a matched filter to each

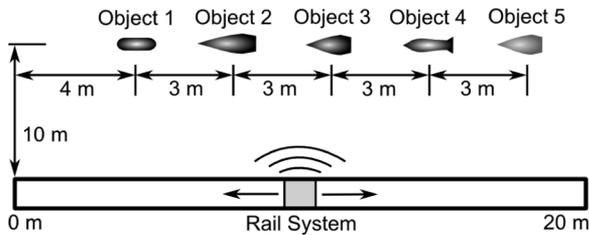


Fig. 4. Ex10 pond target field layout.

sonar return, weighting the result using a reference signal to attenuate the strong bottom return, windowing to remove secondary reflections and other uncorrelated interference, transformation into the frequency domain, and removing the effects of the transmit signal via a pseudoinverse filter.

B. Pond Facility and Collected Data Set

The Naval Surface Warfare Center (NSWC) Ex10 pond facility was designed to collect acoustical sonar data from underwater objects in a relatively controlled and clutter-free environment. Fig. 4 shows the layout of the pond including the relative locations of the rail-mounted sonar system and the objects in the target field. The 20-m rail the sonar system is mounted on is fixed to eliminate platform motion as the sonar moves along its track, thereby increasing coherence between successive pings. The sonar transmit signal is a 6-ms LFM pulse over 0.5–30 kHz with a 10% taper between the leading and trailing edges to minimize ringing in the transmitted signals. Sonar backscatter is received with six hydrophone elements that are arranged in a linear array that is normal to the seafloor, though only data collected by the hydrophone closest to the seafloor are used in this study to generate single-hydrophone SAS-like AC images.

Table II contains descriptions for all the objects in the Ex10 target field, though they were not assigned class labels, while Fig. 4 shows the Ex10 target field layout. As can be seen, the target field contains five objects with varying shapes, sizes, and compositions, all of which are located approximately 10 m horizontally from the rail system and are proud on the sandy bottom. The sonar data set used in this study was collected during ten runs through the target field that differed in the orientation of all the objects, with each object having the same orientation for a given run. Nine total object orientations were used, ranging from -80° to $+80^\circ$ in 20° increments, where two runs used a 0° object orientation, which designates a configuration where the objects' major axes are parallel to the rail system. To collect a single run of data, consisting of 769 pings, the sonar platform moved along the fixed rail in increments of 0.025 m, transmitting and receiving once for each fixed position. The acoustical data were sampled at 1 MHz and the sonar platform was tilted at a fixed 20° grazing angle for all runs (angle of the sonar main response axis with respect to the horizontal plane).

IV. RESULTS

This section presents the results of applying SAS-like AC processing methods to the data sets described in the previous section. First, visual comparisons are made between the images generated by the proposed methods and equivalent conventional

TABLE II
DESCRIPTION FOR EACH OBJECT IN THE EX10 POND DATA SET

Label	Description
Object 1	Small Aluminum Cylinder with Notch
Object 2	De-militarized 152 mm TP-T Round
Object 3	100 mm Solid Steel (SS) Rocket Round
Object 4	Inert 81 mm Mortar (filled with cement)
Object 5	100 mm Machined Aluminum (Al) Rocket Round

SAS processing methods. In particular, the along-track localization capability of SAS-like processing is demonstrated as well as its ability to generate AC patterns that differ between object types. It is also shown that single-hydrophone SAS-like AC images can be used for localizing objects in two dimensions. Next, a simple classifier is introduced that can be applied to SAS-like AC images (multiple-hydrophone case only) representing entire runs of sonar data, effectively performing simultaneous target detection and classification. Finally, the performance of this classifier, and hence, the utility of the AC information provided by the SAS-like images, is demonstrated.

A. Image Comparison

1) *Multiple-Hydrophone Data:* The conventional delay-and-sum SAS processing method in [3] will be used as a reference for evaluating the characteristics of images produced using the multiple-hydrophone SAS-like AC processing in Section II-A. These methods add value to different aspects of the problem of characterizing underwater objects, and hence, their comparison not only verifies that the proposed SAS-like AC method supplies accurate along-track localization capabilities as indicated by a proven original SAS method, but also highlights the advantages and disadvantages of each method for different applications. For conventional delay-and-sum SAS processing in the time domain [3], coherently integrating data from multiple hydrophone elements and pings involves referencing samples from the different time series that correspond to the same focal point or spatial location on the seafloor. A weighted sum of these samples is then calculated to assign an intensity value to the pixel corresponding to the focal point that the referenced samples capture. Although using data from multiple pings and hydrophones to generate each pixel produces a high-resolution image, it carries the burden of high computational complexity owing to the necessity of considering multiple sources of platform motion and near-field propagation to reference the correct samples. This information is often obtained by using an inertial navigation system on a vehicle that records information about its exact position and orientation at each ping.

To compare SAS-like AC processing with conventional delay-and-sum SAS processing, images corresponding to identical areas of the seafloor were generated using each method. Entire runs made by the wing BOSS over the target field described in Section III-A provided the data used for this comparison. These runs contain a variety of target and nontarget objects to obtain an indication of how each algorithm renders different object types, and to demonstrate the utility of the AC information provided by the SAS-like AC images. To

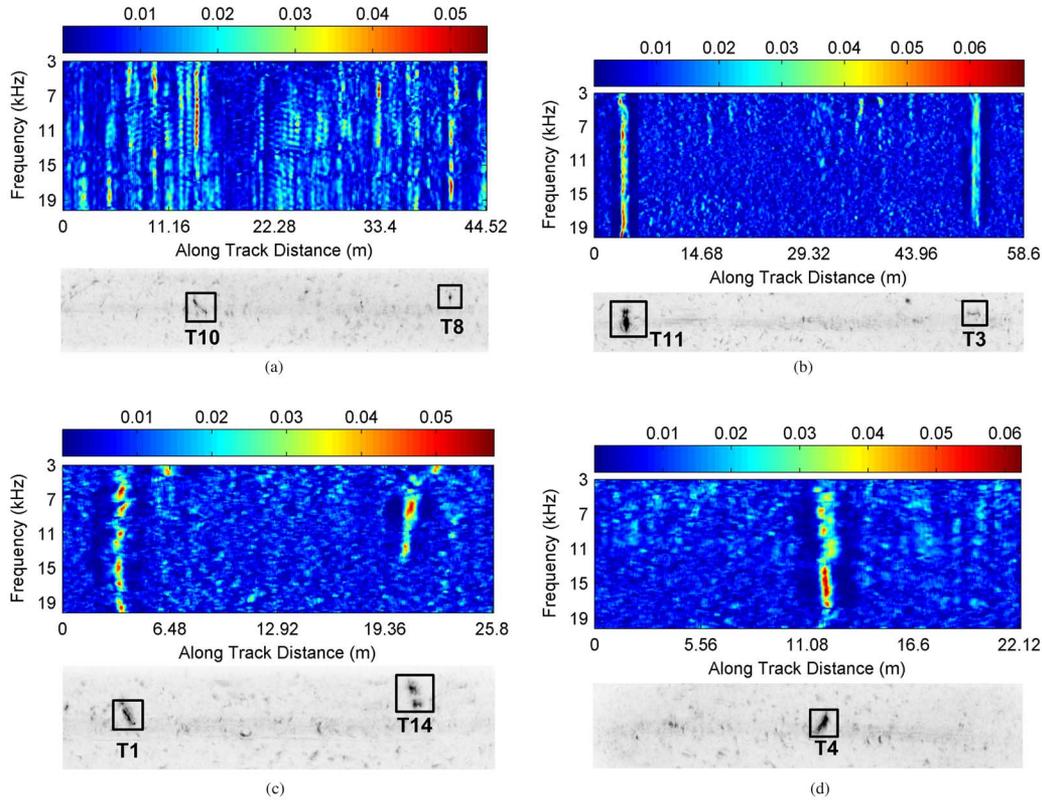


Fig. 5. Comparison of delay-and-sum SAS and SAS-like AC images for different runs through the target field. (a) Run 1c. (b) Run 2d. (c) Run 2i. (d) Run T4e.

ensure a fair comparison, a 71-ping window is used to generate images using both methods. The two input channels of the CCA process are formed using data from two linear subarrays (each with 20 hydrophones) on the wing BOSS sonar platform, with one subarray for each wing of the BOSS. The spacing between the two subarrays is approximately 47 cm as measured by the distance between the hydrophones on the inside of each wing, and the interelement spacing on each subarray is approximately 5.1 cm. A SAS-like AC image is formed by averaging over $Q = 5$ pings to yield the first canonical correlation estimate (see Section II-C) corresponding to the center of the 71-ping window for every frequency sample and ping. Pixels are generated for each frequency sample in the range of 3–19 kHz, with a resolution of 10.5 Hz.

Fig. 5(a)–(d) shows images generated using conventional delay-and-sum SAS processing and SAS-like AC processing for multiple runs through the target field. The top image in each of these figures shows the results of applying SAS-like AC processing to the entire run of data. The colorbar above each SAS-like AC image indicates the mapping between the color of a pixel displayed in the image and the value calculated for $A(f, p)$, which is the magnitude of the estimated dominant canonical correlation for frequency f and ping p [see (14)]. The bottom image in these figures gives the results of applying conventional delay-and-sum SAS processing to the entire run of data. Each pixel in the across-track and along-track plane of this image represents the maximum received signal strength at the corresponding focal point, with lighter pixels representing a weaker response, and darker pixels representing a stronger

response. Each object is outlined with a square in the conventional SAS images to show their locations.

Comparing both images in each of the Fig. 5(a)–(d) demonstrates that a strong response in the conventional SAS image is often paired with a larger coherence value in the same along-track location (ping) in some portion of the spectrum of the SAS-like AC image. Thus, the along-track location of the detected objects can be determined using either image in most cases. Clutter that has unknown characteristics can still be seen in both images, though it appears more clearly in the conventional SAS image since its across-track location and dimensions are revealed. Inspection of the SAS-like AC images shows that, for pings off marked objects, the frequency bands that have prominent coherence traces can reveal useful information about the object's characteristics. Generally, bomb-shaped markers have strong coherence in the mid-to-high frequency ranges. However, the complete behavior of these AC patterns cannot be characterized just by these few images, as they are highly dependent on the range and grazing angle and at which the object was observed, in addition to the object's shape and composition. Additionally, as indicated by other object types that produce strong AC traces in the high-frequency regions, such as T8 in Fig. 5(a) and T11 in Fig. 5(b), classifying an object is not as straightforward as analyzing the strength of the AC traces in isolated bands, but rather, such analysis should incorporate the overall coherence patterns present in all frequency bands. The classification framework introduced in Section IV-B is based on utilizing all available AC information for target versus nontarget discrimination.

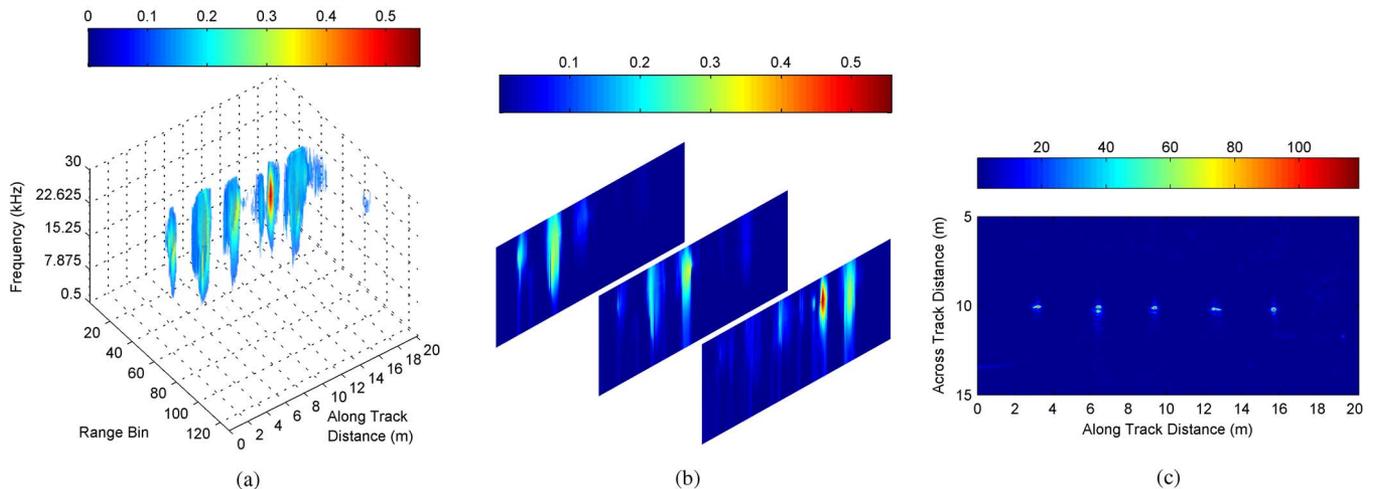


Fig. 6. SAS images for run 185. Major axes of objects are parallel to the rail with their noses pointed towards the sonar starting position (designated 0°). (a) SAS-like AC 3-D solid representation. (b) SAS-like AC range bin slices 53, 54, and 56. (c) Omega- k SAS.

In contrast to SAS-like AC images, conventional SAS images are more suited to characterizing an object based upon shape rather than other characteristics. Indeed, the shapes of the objects are depicted relatively clearly in conventional SAS images, as indicated primarily by the cylindrical appearances of most objects that have this shape. Unfortunately, for this data set, some target objects have the same dimensions as nontarget objects, and hence, the conventional SAS images do not provide enough information for object classification in this case. As mentioned, while the classification ability of the SAS-like AC images may be difficult to achieve visually from these four images alone, their entire AC information does indeed prove to be adequate for this task, as indicated by the results in Section IV-C. Also note that the AC information provided by SAS-like AC images generated using the proposed method may also implicitly provide shape-dependent information, since the shape of an object undoubtedly affects its coherence properties.

2) *Single-Hydrophone Data*: This section presents the results of applying the single-hydrophone SAS-like AC processing method to the Ex10 pond data described in Section III-B. An omega- k beamformer is used to generate SAS images that will be compared with those produced using SAS-like AC processing. The omega- k SAS processing method used here is an adaptation of the omega- k method of synthetic aperture radar (SAR) processing found in [13], which is described in [14], and has been modified by the NSWCC. This method involves deconvolving the aperture amplitude effect in the frequency domain, performing phase multiplication, applying an inverse Stolt mapping, and windowing the result. Using this method, each pixel in a SAS image is generated using all the available ping data in a run collected by a single hydrophone, which produces a high-resolution 2-D image from which objects can be localized in the across-track and along-track dimensions.

To compare SAS-like AC processing with omega- k SAS processing, images corresponding to identical areas of the seafloor were generated using each method. Entire runs over the Ex10 pond target field provided the data used for this com-

parison. Note that, since the images produced by the omega- k and single-hydrophone SAS-like AC methods are drastically different, their comparison is aimed at showing the localization capabilities of the latter as indicated by a proven SAS method. The SAS-like AC images were generated using a synthetic aperture consisting of $P = 3$ hydrophones (data from three different pings) per subarray with a spacing of $J = 6$ pings between the two subarrays. The synthetic subarray size was chosen to be small enough to avoid data poverty that can be introduced by using large dimensional channels for CCA [10]. The subarray separation was chosen to be large enough to ensure low coherence between seafloor clutter captured by each subarray. To provide the necessary realizations of each channel for forming the covariance matrices used in the CCA process the impulse response of the object and bottom at each ping was divided into $L = 128$ different segments, which also allowed the generated 3-D images to contain 128 different range bins for displaying across-track localization information (since the sonar was side-looking). As described in Section II-C, for a single-hydrophone SAS-like AC image, the pixel intensity at each point in a single slice (range bin) of the frequency-ping (i.e., f versus p) plane is the contribution of coherence associated with that range cell segment to the estimate of the dominant canonical correlation. Pixels are generated using (18) for each frequency f (within the 29.5-kHz bandwidth), ping p , and range cell segment l .

Figs. 6–8 show images generated using SAS-like AC processing [images (a) and (b) in each figure] and omega- k SAS processing [image (c) in each figure] for several different runs through the target field. Image (a) in each figure shows a representation of the SAS-like AC volumetric data $B(f, p, l)$, where all voxels satisfying $B(f, p, l) \geq 0.1$ are displayed with a certain transparency. Voxels with an intensity smaller than 0.1 are not displayed in these images so that areas with high levels of coherence, which often represent objects, are displayed more prominently than areas of low coherence, which often represent empty bottom. Image (b) in each figure displays range bin slices of the 3-D SAS-like AC image that contain the largest AC traces

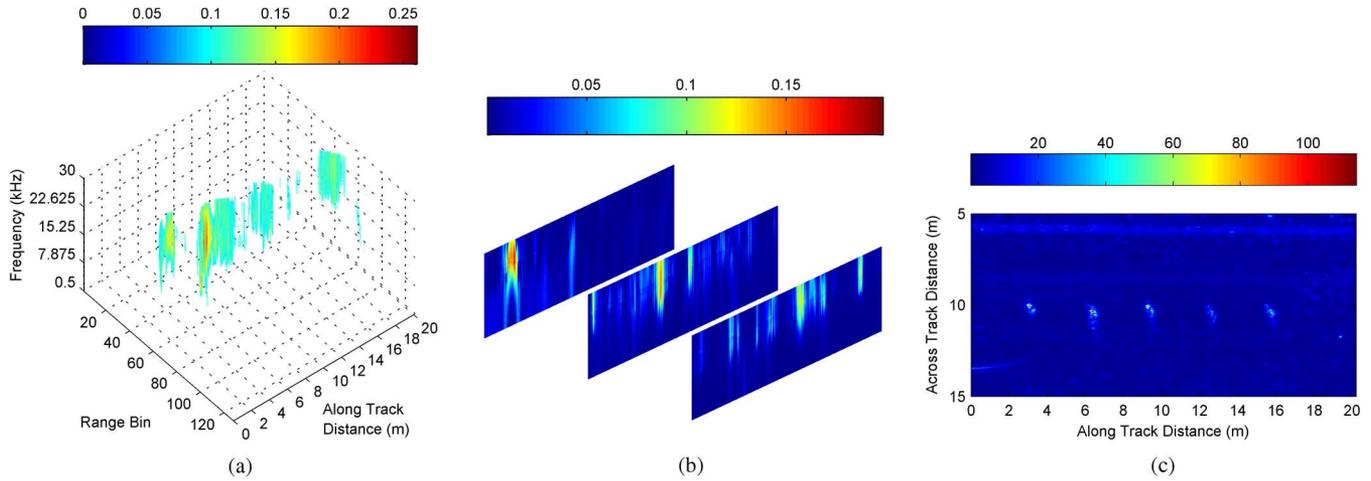


Fig. 7. SAS images for run 191. Major axes of all objects are 40° off parallel with their noses pointed towards the rail (designated -40°). (a) SAS-like AC 3-D solid representation. (b) SAS-like AC range bin slices 52, 63, and 69. (c) Omega-*k* SAS.

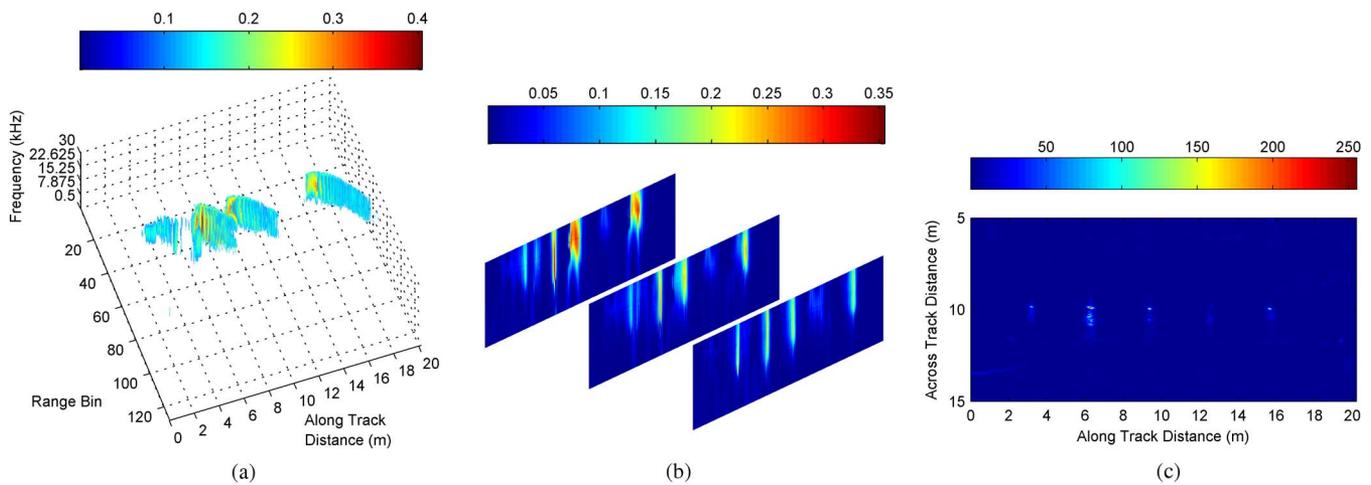


Fig. 8. SAS images for run 203. Major axes of all objects are 80° off parallel with their blunt ends pointed towards the rail (designated +80°). (a) SAS-like AC 3-D solid representation. (b) SAS-like AC range bin slices 50, 51, and 53. (c) Omega-*k* SAS.

(i.e., all pings at a given range are shown for each slice), and are intended to show the detailed AC patterns of labeled objects. Generating ping slice images that show coherence values for all range bins at a given ping is also possible, but not done here. It is important to note that neither the slice images nor the solid image in each case can provide a fully adequate representation of the volumetric data owing to the inherent restrictions of displaying the data on a 2-D plane. In the case of the slice images, only three are shown for each run, and thus, some important AC information present in other slices is sometimes not shown. In the case of the solid images, a single view often cannot accurately portray the location of the stronger AC traces.

Focusing first on Fig. 6, which shows SAS images for a run where all object have their major axes parallel to the rail, it can be seen that a strong response in the omega-*k* SAS image is often paired with a larger coherence contribution in the same location in some portion of the spectrum of the SAS-like AC image. The accuracy of the along-track localization information offered by the SAS-like AC images can be verified by noting the five distinct areas of strong AC traces found in equispaced intervals in the along-track direction, which should be the case

as indicated by the target field layout in Fig. 4 and omega-*k* image in Fig. 6(c). The across-track location of the objects also appears to be consistent between the SAS-like AC image and omega-*k* image for run 185, since all objects appear in nearly the same across-track position, which is in the center of both images. Fig. 6(b) indicates a slight variation in the range bins where the strongest AC traces occur for the different objects, but nonetheless, these range bins are at least adjacent to each other.

The frequency distribution information for each object in run 185, shown in Fig. 6(a) and (b), again demonstrates the ability of the SAS-like AC method to produce AC patterns that vary depending on the type of object captured. Comparing these SAS-like AC images to the target field layout in Fig. 4 it can be seen that object 2 (152-mm TP-T round), object 3 (100-mm SS rocket round), and object 5 (100-mm Al Rocket Round) have similar AC distributions that extend from the low-to-mid frequency range to the high-frequency range, which makes sense since these objects have similar shapes. On the other hand, object 1 (small cylinder) has stronger AC traces mostly in the high-frequency range, and object 4 (mortar) has stronger traces

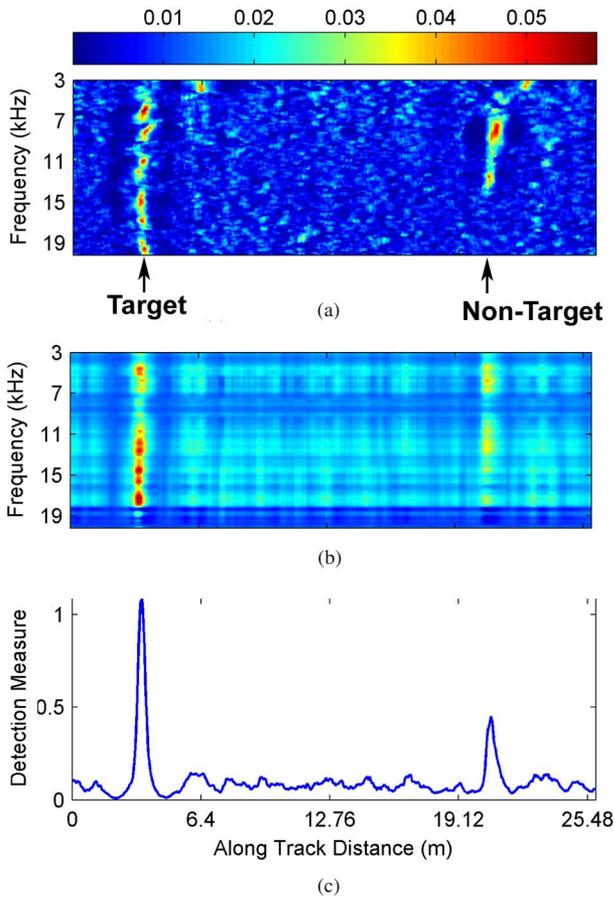


Fig. 9. Example of the detection process using a SAS-like AC image. (a) Original SAS-like image. (b) Projected SAS-like image. (c) Detection measure.

in the midfrequency range and very strong traces in the high-frequency range. As in the multiple-hydrophone case, the ω - k SAS image in Fig. 6(c) appears to lack the ability to convey an adequate amount of information for determining an object's identity since all objects essentially appear as small, high-intensity pixel clusters. However, compared with the SAS-like AC image, the ω - k image seems to provide more refined localization in both dimensions. This is due to the precise motion compensation employed by this method.

The SAS-like AC images for runs 191 and 203, shown in Figs. 7 and 8, respectively, can be used to see how object orientation affects the AC patterns. Note that, for runs 191 and 203, all objects were oriented -40° and $+80^\circ$, respectively, relative to the parallel orientation employed in run 185. The effects of these reorientations are dramatic, though objects can still be localized for the most part. Specifically, looking at the SAS-like AC images for run 191 in Fig. 7, all objects except object 4 can be seen prominently at some range bin in the proper along-track location, but the representative AC traces appear weaker than those in run 185. Very weak AC traces can be seen for object 4, which matches the weak signatures of this object for the ω - k image in Fig. 7(c). The cause of these weaker signatures is likely the small diameter of this object relative to the others, which makes it difficult for the sonar system to capture it considering the orientation used for run 191. Though it may be

difficult to discern when restricted to a single volumetric view of the 3-D SAS image, the AC distributions for each object in run 191 are largely the same as those witnessed for objects in run 185. Different alterations of the AC patterns can be seen for run 203 in Fig. 8(a) and (b), though the AC traces appear to be more prominent than those for run 191. For this run, the AC traces of object 4 are missing completely, though this object can barely be seen in the ω - k image for this run as well [see Fig. 8(c)].

Based on the above results, it is clear that both conventional SAS processing and SAS-like AC processing methods generate images that offer valuable information about objects that are buried beneath the seafloor or proud on the bottom. Applications requiring more accurate across-track localization, determination of an object's dimensions, or high-resolution depictions of subtle contents of the seafloor would benefit from using SAS images generated using conventional methods (delay-and-sum or ω - k). Though the classification utility of the AC information provided by SAS-like processing will be demonstrated in Section IV-C, for now the ability of this method to generate images offering fine along-track and coarse across-track localization has been demonstrated. As anticipated, changing the orientation of an object relative to the sonar affects the prominence as well as distribution of the generated AC patterns.

B. Subspace Projection for Object Classification

This section describes a simple but effective method for classifying pings in a 2-D SAS-like AC image as either target or nontarget, taking advantage of the AC patterns typically observed for pings that capture target objects. This method is applicable to either a SAS-like AC image generated using the multiple-hydrophone method, or a single slice of the image generated using the single-hydrophone method. This framework individually classifies SAS-like AC feature vectors, each of which is a column vector of pixels in a SAS-like AC image at different frequencies and at a specific ping, i.e., a single column of $A(f, p)$ in (14) at a given ping p , or a specific cross section of two planes for a fixed ping p and fixed range l of the volumetric image $B(f, p, l)$ in (18). Classification of a single ping is simply performed by finding the magnitude of the portion of its associated AC feature vector that lies within a "target subspace." In essence, such a measure is able to quickly determine how similar a new testing AC feature vector is to training AC feature vectors that are known to represent pings off target objects. This is because such training AC feature vectors are used to form this target subspace as outlined below.

Let us define $S = [s_1 s_2 \cdots s_k \cdots]$ as a training data matrix containing different SAS-like AC training feature vectors (or patterns) s_k 's, representing what are known to be pings off a variety of target objects at different orientations, elevations, and grazing angles. The training features s_k 's should ideally come from many different runs made by a UUV and be representative of objects with different shapes, compositions, and burial conditions. Application of SVD to data matrix S yields

$$S = \Psi \Sigma \Phi^T \quad (19)$$

where $\Psi = [\psi_1 \psi_2 \cdots]$ and $\Phi = [\phi_1 \phi_2 \cdots]$ contain left and right singular vectors (as columns), respectively, and Σ is a diagonal matrix of singular values. The well-known Eckart-Young

theorem [16] states that the dominant left singular vectors in Ψ can be used to construct an optimal low-rank approximation of S (minimum Frobenius norm compared with other approximations of the same rank). Therefore, assuming the singular values that are the diagonal elements of Σ appear in descending order of magnitude, the first K columns of Ψ , corresponding to K dominant singular values, that comprise the matrix $\Psi^{(K)} = [\psi_1 \cdots \psi_K]$, are used as basis vectors to span the target subspace. If S was formed using appropriately diverse target sample AC feature vectors, and K is chosen to be adequately large, then we can expect that AC feature vectors representing target objects lie mostly within this target subspace. For the classifier described below, this target subspace, which is spanned by the columns of $\Psi^{(K)}$, is characterized by its orthogonal projection matrix $P_\Psi = \Psi^{(K)}\Psi^{(K)T}$.

Using the projection matrix, the classifier used in this study evaluates the unknown AC pattern for each ping of interest p by projecting it onto the target subspace using

$$d_p = \begin{array}{l} \text{Label target} \\ \|\!|P_\Psi \mathbf{s}_p\!\|_2 \geq \\ \text{Label nontarget} \end{array} \eta \quad (20)$$

where $P_\Psi \mathbf{s}_p$ is the projected AC feature vector, d_p is the measure (magnitude of the projected AC pattern) used to determine the class label for ping p , and η is a threshold whose value can be determined by the examining the values of d_p calculated for training AC feature vectors representing various target and nontarget objects. Essentially, (20) implements a matched subspace detector/classifier [17] at each ping of interest, and states that ping p should be labeled as target when its corresponding AC pattern \mathbf{s}_p is very similar to the training target feature vectors \mathbf{s}_k 's that are the columns of S . This a consequence of the fact that

$$P_\Psi = \arg \min_{\{P: \text{rank}(P)=K\}} \|\!|S - PS\!\|_F^2 \quad (21)$$

where $\|\cdot\|_F^2$ means Frobenius norm. Apart from its simplicity, one of the main benefits of this classification method is the fact that it can effectively deal with the large dimensionality of AC patterns; e.g., for the BOSS data set each \mathbf{s}_p has 1639 elements. In such conditions, many conventional classifiers, such as a neural network that is based on learning the distribution of the input pattern, would be impractical or extremely difficult to use.

Fig. 9 demonstrates the idea behind the proposed classification method, where Fig. 9(a) shows the original SAS-like AC image for a single run of data, Fig. 9(b) shows the SAS-like patterns for this run after projection onto the target subspace, and Fig. 9(c) shows the measure d_p for each ping in the run. As can be seen, when an AC pattern \mathbf{s}_p represents a ping off a nontarget object, the operation $P_\Psi \mathbf{s}_p$ attenuates AC information that is dissimilar to target objects (not frequently seen in the columns of S) and overall decreases its magnitude more significantly than the target pings. Additionally, since pings off empty bottom produce weak coherence patterns, an AC feature vector \mathbf{s}_p representing empty bottom leads to a small d_p for such a ping.

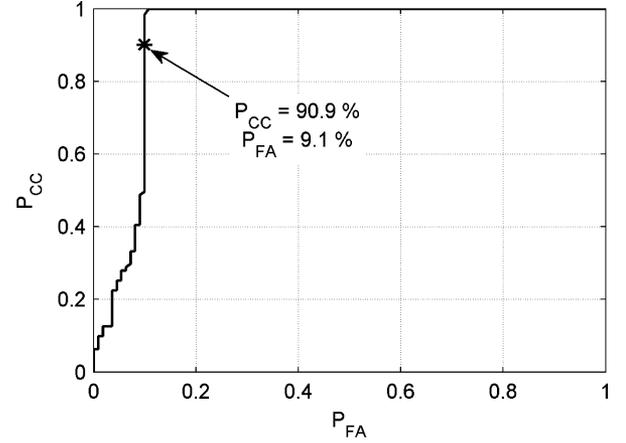


Fig. 10. ROC curve for the testing set.

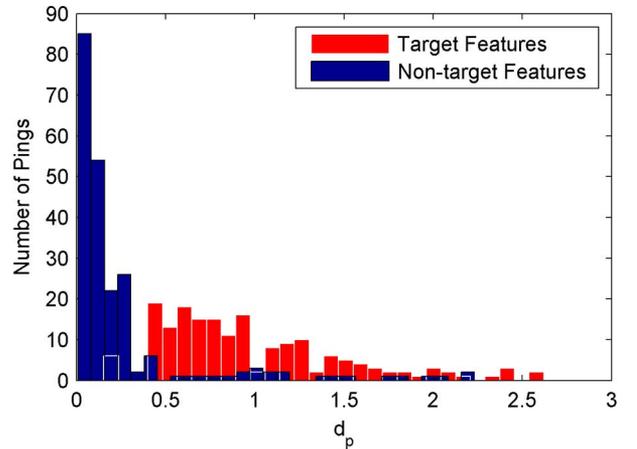


Fig. 11. Distribution of d_p for target and nontarget pings.

This implies that this method can be used both as a detector (object versus background) and a classifier (target versus nontarget) at the same time.

The flexibility of this classification method lies in the fact that it does not require objects in specific classes to have certain predefined characteristics, such as restrictions on their shapes or sizes. As mentioned, to achieve adequate classification performance, it is only necessary to form a target subspace (characterized by P_Ψ) using a data matrix S containing samples that are representative of the various AC patterns typically seen for target objects. Therefore, even if objects in different classes have some similarities, adequate classification results can still be obtained as long as there are AC features that the target group of objects has in common. This is especially important for classification of objects in the data set described in Section III-A, as many objects with opposite class labels have similar sizes with different compositions or, more rarely, similar compositions with different dimensions. Because of these characteristics, in Section IV-C, we test the effectiveness of the proposed SAS-like AC processing and the simple classification method in this section on the multiple-hydrophone data set.

C. Classification Results

This section presents the results of applying the subspace projection classifier described in Section IV-B to pings in the

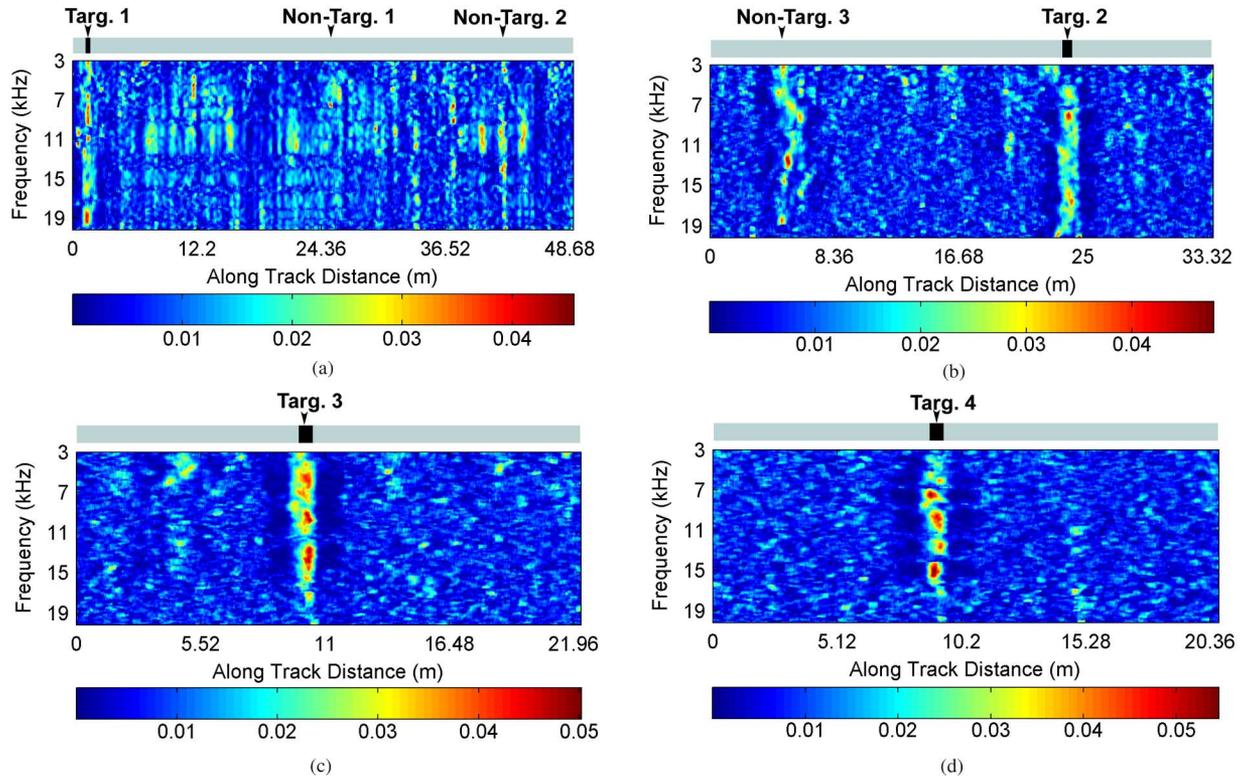


Fig. 12. Examples of classifying entire runs of sonar data using SAS-like AC images. (a) Testing run 2b. (b) Testing run 2h. (c) Training run T4c. (d) Training run T10i.

BOSS data set. First, the SAS-like AC patterns associated with pings off objects will be evaluated to test the proposed method's ability to generate patterns that can be used to distinguish between target and nontarget objects. The results of applying the subspace projection detector/classifier to entire runs of data (all pings in a run) will then be presented to demonstrate this method's ability to highlight pings where target objects are present in the absence of any prior knowledge concerning their locations.

To form the target subspace, $K = 2$ basis vectors were extracted from a data matrix S , consisting of SAS-like AC feature vectors representing pings off target objects captured during seven different runs from the groups of runs made by the UUV that were centered over specific objects. More specifically, these training AC feature vectors represent 68 pings off various target objects in nine different runs. The resulting target basis vectors completely characterize the proposed classifier by forming the matrix P_{Ψ} used in (20). The threshold η is selected based on a comparison of the distributions of d_p for target and nontarget pings in the training runs. The testing set, used to evaluate the ability of the classifier using SAS-like AC patterns to discriminate between different object types, contains pings off target objects captured during 12 northeast–southwest runs where the UUV traversed the entire target field. The target pings in the testing set offer a unique perspective of these objects since different grazing angles and vehicle trajectories were used to collect these pings. This testing set also contains pings from all nontarget objects captured during all 19 of the aforementioned runs.

Fig. 10 shows the receiver operating characteristic (ROC) curve generated by applying the proposed classifier to all SAS-

like AC feature vectors in the testing set. Note that to generate this ROC curve an equal number of AC patterns from each class is used (to avoid bias); a condition that was met by discarding a random selection of the relatively more plentiful nontarget AC patterns. As can be seen from Fig. 10, we have $P_{CC} = 90.9\%$ and $P_{FA} = 9.1\%$ at the knee point (where $P_{CC} + P_{FA} = 1$) of the ROC curve. The area under the ROC curve is 0.925, which corresponds to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one. The shape of the ROC curve demonstrates the fact that, once a certain threshold is reached, $P_{CC} = 100\%$ can be achieved with only a 10.8% false alarm rate. This is because the proposed classification measure in (20) was larger for the majority of pings off target objects when compared with pings off nontarget objects, and hence, it provided excellent class separability. Fig. 11 further demonstrates the class separability offered by the use of d_p , as it shows the distributions of d_p for all target and nontarget pings in the BOSS data set. As can be seen from this figure, the majority of nontarget pings have a much smaller associated d_p value when compared with those of target pings. The sporadic instances of overlap of these distributions that cause a slight decrease in classification performance are discussed at the end of this section.

Applying the proposed classification framework to entire runs of data involves projecting the SAS-like AC feature vector representing each ping in the run onto the target subspace and calculating $d_p, \forall p$ in the run. The values of d_p that exceed the fixed threshold of $\eta = 0.4$ are considered target but are otherwise considered to be nontarget. Fig. 12(a)–(d) shows the classification results obtained by applying the proposed classifica-

tion framework to all the corresponding extracted SAS-like AC feature vectors from four separate runs through the target field introduced in Section III-A. Shown at the top of these figures are detection/classification strips that illustrate at which sonar pings AC patterns were declared as target (detected and classified). In these strips, the black-colored portions represent the pings at which the system has declared a target object. To indicate locations where the detection/classification strips would ideally declare the presence of a target object, all object locations are labeled above the strips for reference.

Fig. 12(a)–(d) demonstrates the ability of the SAS-like AC method to produce features that can be used to discriminate between sonar returns off target objects from those off nontarget objects and empty bottom, as the detection/classification strips indicate the presence of target objects only where they are indeed present. When the simple subspace projection-based classification system in Section IV-B was applied to all pings from every run used in this study it was found that, at a correct classification rate of 96.6% on target objects, where only some of the pings off one target object in a single run were incorrectly assigned the nontarget class label, a false alarm rate of only 1.29% was achieved. Out of this low false alarm rate only 11.7% were attributed to nontarget object pings while 88.3% of them were attributed to the relatively more plentiful empty bottom returns [none occurred for the runs displayed in Fig. 12(a)–(d)].

At the chosen threshold of $\eta = 0.4$ the main causes of misclassification were one collection of pings off one target object having low d_p , as well as some pings off two nontarget objects having relatively high d_p . In the former case, the SAS-like AC features extracted from the target pings did contain high levels of coherence in a frequency band that is typical for target pings, but this coherence was isolated to a very small frequency band, leading to an overall lower value of d_p for these pings. The misclassified target object does not appear very prominently in the conventional SAS image generated for this run either (not shown), thus it is likely that the signatures of this object were not adequately captured by the sonar for this particular run. The misclassified pings off nontarget objects were because these pings exhibited high levels of coherence in frequency bands that were typical for target pings. This display of target-like behavior may be attributed to the fact that the misclassified nontarget objects have the same dimensions as some target objects, despite having different compositions. Since the shape of an object is one of the factors that undoubtedly affects its SAS-like AC patterns, misclassification in this case was understandable. It may be possible to avoid such errors by using a larger number of training features to form the target subspace, thus more accurately characterizing the relevant target AC features.

Based on the above results, it is clear that the proposed SAS-like AC processing is capable of generating images that offer valuable information about objects that are buried beneath the seafloor or proud on the bottom. In particular, the proposed method has shown its promise in generating images that convey useful information about an object's type, in addition to its location. Furthermore, creating SAS-like AC images does not require direct motion compensation, and consequently does not need to use information from the inertial navigation systems on the vehicle. This reduces the overall cost of the SAS system on

UUVs, which could be an important factor if smaller size and lower cost UUVs are to be deployed in large numbers.

V. CONCLUSION

This study developed two different methods of applying a coherence analysis framework to broadband sonar data to generate SAS-like AC images that can be used for underwater object detection, localization, and classification without the use of elaborate platform motion estimation and compensation processes. These two methods are applicable to data collected with either multiple hydrophones, or single-hydrophone sonar. Both methods generate an image by forming two data channels, representing different real or synthetic hydrophone subarrays, and extracting the canonical correlations between these channels. Estimates of the dominant canonical correlation are then used to calculate the intensity of a pixel in the ping-frequency plane. It was also shown that, by averaging over frequency samples from different range cell segments to form the necessary covariance matrices, the resulting SAS-like AC images can be used to localize objects in two dimensions.

Comparing the SAS images generated using conventional methods and SAS-like AC methods, it was observed that they both offer valuable information about objects that are buried beneath the seafloor or proud on the bottom. Specifically, conventional (delay-and-sum or omega- k) SAS processing has reaffirmed its effectiveness in generating images for object detection and accurate localization, whereas the proposed SAS-like AC processing has shown its promise in generating images that can be used for composition-based object classification as well as accurate along-track localization and, in the case of the single-hydrophone method, possible coarse across-track localization. A method for classifying pings based on projecting the SAS-like AC feature vectors onto a target subspace was also introduced. Classification results demonstrated that the AC information produced by the proposed method is indeed useful for object class discrimination.

APPENDIX

This Appendix reviews CCA, which is a method used for analyzing the linear dependence (or coherence) between two data channels. CCA decomposes the linear dependence between the original channels into the linear dependence between the canonical coordinates of the channels where this linear dependence is easily determined by the corresponding canonical correlations. The material presented here and much of the language and terminology are drawn from [12].

Consider two zero mean random vectors $\mathbf{x} \in \mathbb{C}^m$ and $\mathbf{y} \in \mathbb{C}^n$ where it is assumed that $m \leq n$. The composite data vector $\mathbf{z} = [\mathbf{x}^T \mathbf{y}^T]^T \in \mathbb{C}^{(m+n)}$ has a covariance matrix

$$R_{\mathbf{z}\mathbf{z}} = E[\mathbf{z}\mathbf{z}^H] = \begin{bmatrix} R_{\mathbf{x}\mathbf{x}} & R_{\mathbf{x}\mathbf{y}} \\ R_{\mathbf{y}\mathbf{x}} & R_{\mathbf{y}\mathbf{y}} \end{bmatrix} \quad (22)$$

where H is the Hermitian operator, $R_{\mathbf{x}\mathbf{x}}$ and $R_{\mathbf{y}\mathbf{y}}$ are the covariance matrices of \mathbf{x} and \mathbf{y} , respectively, and $R_{\mathbf{x}\mathbf{y}}$ is the cross-covariance matrix between \mathbf{x} and \mathbf{y} . The coherence matrix is then formed using $C = R_{\mathbf{x}\mathbf{x}}^{-1/2} R_{\mathbf{x}\mathbf{y}} R_{\mathbf{y}\mathbf{y}}^{-H/2}$, where we have

$R_{\mathbf{x}\mathbf{x}}^{-1/2} R_{\mathbf{x}\mathbf{x}} R_{\mathbf{x}\mathbf{x}}^{-H/2} = I$, and $R_{\mathbf{x}\mathbf{x}}^{1/2} R_{\mathbf{x}\mathbf{x}}^{H/2} = R_{\mathbf{x}\mathbf{x}}$. The SVD of the coherence matrix C yields

$$C = FKG^H \text{ and } F^H CG = K \quad (23)$$

where $F \in \mathbb{C}^{m \times m}$ and $G \in \mathbb{C}^{n \times n}$ are orthogonal matrices and $K = [K(m) \ \mathbf{0}] \in \mathbb{R}^{m \times n}$ is a singular value matrix, with $K(m) = \text{diag}[k_1, k_2, \dots, k_m]$ and $1 \geq k_1 \geq k_2 \geq \dots \geq k_m > 0$.

Using the orthogonal matrices F and G , the canonical coordinates of \mathbf{x} and \mathbf{y} , denoted by $\mathbf{u} = [u_i]_{i=1}^m \in \mathbb{C}^m$ and $\mathbf{v} = [v_i]_{i=1}^n \in \mathbb{C}^n$, respectively, are then defined as

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} F^H & \mathbf{0} \\ \mathbf{0} & G^H \end{bmatrix} \begin{bmatrix} R_{\mathbf{x}\mathbf{x}}^{-1/2} & \mathbf{0} \\ \mathbf{0} & R_{\mathbf{y}\mathbf{y}}^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}. \quad (24)$$

The vectors \mathbf{u} and \mathbf{v} share the cross-correlation matrix $R_{\mathbf{u}\mathbf{v}} = E[\mathbf{u}\mathbf{v}^H] = K$ while $R_{\mathbf{u}\mathbf{u}} = E[\mathbf{u}\mathbf{u}^H] = I(m)$ and $R_{\mathbf{v}\mathbf{v}} = E[\mathbf{v}\mathbf{v}^H] = I(n)$. The cross-correlation matrix K is called the canonical correlation matrix of canonical correlations k_i , $i = 1, \dots, m$. Thus, the canonical correlations measure the correlations between pairs of corresponding canonical coordinates. That is, $E[u_i v_j] = k_i \delta_{ij}$; $i \in [1, m]$, $j \in [1, n]$, with δ_{ij} being the Kronecker delta. The canonical correlations k_i are also the singular values of the coherence matrix C , and hence can be used as a measure of coherence between the data channels \mathbf{x} and \mathbf{y} .

It has been shown [12] that the linear dependence between the two data channels \mathbf{x} and \mathbf{y} can be measured by

$$L = \det(I - KK^T) = \prod_{i=1}^m (1 - k_i^2), \quad 0 \leq L \leq 1. \quad (25)$$

The measure L takes the value 0 if there is perfect linear dependence between the two channels \mathbf{x} and \mathbf{y} , and it takes the value 1 if \mathbf{x} and \mathbf{y} are linearly independent. The i th term of the product on the right-hand side of (25), i.e., $(1 - k_i^2)$, measures the linear dependence between the i th canonical coordinate of \mathbf{x} and \mathbf{y} . This implies that the linear dependence between \mathbf{x} and \mathbf{y} is decomposed into the linear dependence between their canonical coordinates, which is easily measured by their canonical correlations k_i 's. Correspondingly, the coherence measure between the elements of \mathbf{x} and \mathbf{y} is defined as $H = 1 - L$; $0 \leq H \leq 1$. Now, if \mathbf{x} and \mathbf{y} are perfectly coherent, $H = 1$, while for the noncoherent case, $H = 0$.

ACKNOWLEDGMENT

The authors would like to thank R. Holtzapple, J. Lopes, and J. D. Tucker at NSWC for providing the data used in this work.

REFERENCES

- [1] M. P. Hayes and P. T. Gough, "Broad-band synthetic aperture sonar," *IEEE J. Ocean. Eng.*, vol. 17, no. 1, pp. 80–94, Jan. 1992.
- [2] D. Cook, J. T. Christoff, and J. E. Fernandez, "Motion compensation of AUV-based synthetic aperture sonar," in *Proc. MTS/IEEE OCEANS Conf.*, Sep. 2003, vol. 4, pp. 2143–2148.
- [3] S. G. Schock, J. Wulf, G. Quentin, and J. Sara, "Synthetic aperture processing of buried object scanning sonar data," in *Proc. MTS/IEEE OCEANS Conf.*, Sep. 2005, vol. 3, pp. 2236–2241.
- [4] S. G. Schock and J. Wulf, "Imaging performance of BOSS using SAS processing," in *Proc. MTS/IEEE OCEANS Conf.*, Sep. 2006, DOI: 10.1109/OCEANS.2006.306896.
- [5] K. Lo, "Adaptive array processing for wide-band active sonars," *IEEE J. Ocean. Eng.*, vol. 29, no. 3, pp. 837–846, Jul. 2004.
- [6] I. Solomon and A. Knight, "Spatial processing of signals received by platform mounted sonar," *IEEE J. Ocean. Eng.*, vol. 27, no. 1, pp. 57–65, Jan. 2002.
- [7] N. Intrator, N. Neretti, and Q. Huynh, "Sonar object discrimination via spectral density analysis," in *Proc. MTS/IEEE OCEANS Conf.*, Nov. 2004, vol. 2, pp. 740–742.
- [8] D. W. Miklovic and M. T. Bird, "Quantitative acoustic color displays for classification with broadband sonars," in *Proc. MTS/IEEE OCEANS Conf.*, Sep. 1998, vol. 1, pp. 592–597.
- [9] R. Duda, P. Hart, and D. G. Stork, *Pattern Classification*. New York: Wiley, 2001, pp. 216–223.
- [10] Y. Zhao, M. R. Azimi-Sadjadi, N. Wachowski, and N. Klausner, "Spatial correlation analysis using canonical correlation decomposition for sparse sonar array processing," in *Proc. IEEE Int. Conf. Syst. Man Cybern.*, Oct. 2009, pp. 2739–2744.
- [11] Y. Zhao, N. Wachowski, and M. R. Azimi-Sadjadi, "Target coherence analysis using canonical correlation decomposition for SAS data," in *Proc. MTS/IEEE OCEANS Conf.*, Oct. 2009, pp. 1–7.
- [12] L. L. Scharf and C. T. Mullis, "Canonical coordinates and the geometry of inference, rate, and capacity," *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 824–831, Mar. 2000.
- [13] M. C. Cobb and J. H. McClellan, "Omega-k quadtree UWB SAR focusing," in *Proc. IEEE Radar Conf.*, May 2001, pp. 311–314.
- [14] D. W. Hawkins, "Synthetic aperture imaging algorithms: With applications to wide-bandwidth sonar," Ph.D. dissertation, Dept. Electr. Electron. Eng., Univ. Canterbury, Christchurch, New Zealand, 1996.
- [15] G. I. Allen, G. Sulzberger, J. T. Bono, J. S. Pray, and T. R. Clem, "Initial evaluation of the new real-time tracking gradiometer designed for small unmanned underwater vehicles," in *Proc. MTS/IEEE OCEANS Conf.*, Sep. 2005, vol. 3, pp. 1956–1962.
- [16] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 11, pp. 211–218, Sep. 1936.
- [17] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2146–2157, Aug. 1994.



Neil Wachowski (S'09) received the B.S. degree in electrical engineering from Michigan Technological University, Houghton, in 2002 and the M.S. degree in electrical engineering from Colorado State University (CSU), Fort Collins, in 2009, with specialization in signal processing. Currently, he is working towards the Ph.D. degree at CSU with specialization in signal processing.

He was a Field Applications Engineer at Texas Instruments, Inc. from 2002 to 2006. His research interests include detection and parameter estimation theory, digital signal/image processing, and system identification.



Mahmood R. Azimi-Sadjadi (S'81–M'81–SM'89) received the M.S. and Ph.D. degrees both in electrical engineering with specialization in digital signal/image processing from the Imperial College of Science and Technology, University of London, London, U.K., in 1978 and 1982, respectively.

Currently, he is a Full Professor at the Electrical and Computer Engineering Department, Colorado State University (CSU), Fort Collins. He is also the Director of the Digital Signal/Image Laboratory, CSU. His main areas of interest include digital signal and image processing, wireless sensor networks, target detection, classification and tracking, adaptive filtering, system identification, and neural networks. His research efforts in these areas resulted in over 250 journal and referenced conference publications. He is the coauthor of the book *Digital Filtering in One and Two Dimensions* (New York: Plenum, 1989).

Prof. Azimi-Sadjadi served as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE TRANSACTIONS ON NEURAL NETWORKS.