

# Fast Target Detection in Synthetic Aperture Sonar Imagery: A New Algorithm and Large-Scale Performance Analysis

David P. Williams

**Abstract**—In this paper, a new unsupervised algorithm for the detection of underwater targets in synthetic aperture sonar (SAS) imagery is proposed. The method capitalizes on the high-quality SAS imagery whose high resolution permits many pixels on target. One particularly novel component of the method also detects sand ripples and estimates their orientation. The overall algorithm is made fast by employing a cascaded architecture and by exploiting integral-image representations. As a result, the approach makes near-real-time detection of proud targets in sonar data onboard an autonomous underwater vehicle (AUV) feasible. No training data are required because the proposed method is adaptively tailored to the environmental characteristics of the sensed data that are collected *in situ*. To validate and assess the performance of the proposed detection algorithm, a large-scale study of SAS images containing various mine-like targets is undertaken. The data were collected with the MUSCLE AUV during six large sea experiments, conducted between 2008 and 2012, in different geographical locations with diverse environmental conditions. The analysis examines detection performance as a function of target type, aspect, range, image quality, seabed environment, and geographical site. To our knowledge, this study—based on nearly 30 000 SAS images collectively covering approximately 160 km<sup>2</sup> of seabed, and involving over 1100 target detection opportunities—represents the most extensive such systematic, quantitative assessment of target detection performance with SAS data to date. The analysis reveals the variables that have the largest impact on target detection performance, namely, image quality and environmental conditions on the seafloor. Ways to exploit the results for adaptive AUV surveys using through-the-sensor data are also suggested.

**Index Terms**—Algorithm, autonomous underwater vehicle (AUV), detection, mine countermeasures (MCM), performance analysis, synthetic aperture sonar (SAS).

## I. INTRODUCTION

**D**URING military campaigns, the deployment of underwater mines by adversaries must be combatted to ensure the safe transport of troops and matériel. In the civilian sector during peacetime, the potential detonation of those same mines left behind after conflicts threatens to disrupt global supply

chains and shipping routes. In both cases, the ability to reliably detect these insidious weapons of war is of paramount importance.

The high-resolution imaging of underwater environments afforded by sonar has proven particularly useful for the detection of proud mines on the seabed. Thanks to advances in marine robot technology, the sonar data used to address this task are very often collected by an autonomous underwater vehicle (AUV). Because of the inherent danger and time-sensitive nature of such mine countermeasure (MCM) missions, the next urgent priority is to embed intelligence in the AUV so that it can immediately react to the data it collects. By adapting its survey route *in situ* and efficiently allocating resources, the AUV can collect the most informative data for the task at hand while simultaneously reducing mission times and ameliorating the problems of “data deluge” and operator fatigue [1]. Doing so, however, requires an algorithm that can perform robust mine detection in near real time onboard an AUV with limited processing capabilities. The existing detection algorithms in the research community suffer from several limitations that preclude this possibility.

The advent of high-resolution synthetic aperture sonar (SAS) [2] means that the task of mine detection is now considerably easier. The excellent resolution afforded by SAS systems directly translates into more pixels on target, the most critical parameter for image-based detection [3]. As a result, a relatively simple detection algorithm that exploits this improved image resolution can be sufficient for achieving good detection performance. This work presents a computationally fast, easily tunable detector design that is intended for, and that capitalizes on, high-quality SAS imagery whose high resolution permits many pixels on target.

Most of the existing mine detection algorithms, which were originally designed for lower resolution sidescan sonar imagery rather than SAS imagery, simply search for highlight-shadow patterns characteristic of mines [4]–[9]. (One notable exception is [10], which was developed specifically for dual-band, high-frequency and broadband, SAS data.) The highlight is the result of the acoustic echo from the mine itself, while the shadow that is cast is due to the geometry between the mine (and specifically its height above the seafloor) and the grazing angle of the transmitted signal. Unfortunately, most of the extant mine detection algorithms suffer from at least one of four major limitations. First, all of these detection algorithms implicitly assume, incorrectly, that the image quality is uniformly excellent across

Manuscript received August 20, 2013; revised November 08, 2013; accepted December 05, 2013.

Associate Editor: T. Cobb.

The author is with the Centre for Maritime Research & Experimentation (CMRE), NATO Science and Technology Organization, La Spezia 19126, Italy (e-mail: williams@cmre.nato.int).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JOE.2013.2294532

the entire image. Second, many detection algorithms do not sufficiently exploit the range-dependent nature of target echoes and shadows that can be predicted based on geometry. Third, most detection algorithms are not tailored to the environmental conditions where the surveying occurs, relying instead on training data from a different site; in addition to the potential data mismatch that this can introduce, using training data also limits the types of targets that can be expected to be detected. Fourth, some detection algorithms, such as [8], rely on relatively slow matched-filtering-type approaches that employ discrete Fourier transforms and preclude the possibility of near-real-time<sup>1</sup> detection. One exception *vis-à-vis* speed is the algorithm presented in [9], which also makes use of integral images, albeit in a fundamentally different manner compared to our proposed approach. The algorithm in [9], which instead suffers from the other aforementioned limitations, merges the detection and classification stages by considering thousands of generic Haar-based features in a brute-force strategy that requires access to a set of training data for learning.

This paper makes two principal contributions. The first contribution is the introduction of a new algorithm for the detection of underwater targets in sonar imagery that overcomes several limitations and makes near-real-time detection onboard an AUV feasible. The second main contribution of this work is a large-scale performance assessment undertaken to validate the proposed algorithm and study the capability of detecting underwater mine-like targets in SAS imagery. The analysis examines detection performance as a function of target type, aspect, range, image quality, and seabed environment by using real SAS data collected during six large sea experiments, conducted between 2008 and 2012, in different geographical locations with diverse environmental conditions. To our knowledge, this study—based on nearly 30 000 SAS images collectively covering approximately 160 km<sup>2</sup> of seabed, and involving over 1100 target detection opportunities—represents the most extensive such systematic, quantitative assessment of target detection performance with SAS data to date. The analysis reveals the variables that have the largest impact on mine detection performance, insight that can further inform the development of intelligent, adaptive, completely autonomous underwater surveys without any “human in the loop.”

Although the primary focus of the work here is on the task of mine detection, the algorithms can also be used for a wide range of other object-detection-related underwater applications, including habitat mapping [11], seabed classification [12], archaeology [13], and pipeline monitoring [14]. Importantly, with trivial modifications, the detection algorithm can be employed for sidescan sonar images, which are still in more common use than SAS images, assuming the sizes of the objects are still large relative to the sonar’s resolution. But the methods proposed are also relevant to fields that employ other sensor modalities for image-processing-based tasks. For example, the fast detection framework can be exploited for ground-based target detection [15], while the ripple detection algorithm would be useful for

oil exploration applications that require the detection of sand dunes [16].

The remainder of this paper is organized as follows. Section II describes the proposed detection algorithm in detail. Section III explains the real, measured sonar data for which detection results are presented in Section IV. A brief discussion of the algorithm and results are provided in Section V, followed by concluding remarks in Section VI.

## II. PROPOSED DETECTION ALGORITHM

### A. Overview

The target-detection algorithm described in detail here is an integral-image-based approach that enables near-real-time detection to be performed onboard a sonar-equipped AUV. Earlier preliminary versions of the algorithm were published (in conference papers) first in [17] and then, after the development of the ripple detection stage, in [18], but still further significant modifications have been made subsequently. Among these changes are the inclusion of a side-echo calculation that helps reduce the false alarm rate, the use of contextual information in a filtering step to reject alarms that collectively constitute an unrealistically high density of contacts, and a new final detection confidence score (based on two novel features) that more accurately orders alarms. The substantial value of the new final detection score will be demonstrated in detail in Section IV. Additionally, this work is important and sufficiently novel because a large-scale performance assessment of the detection algorithm had never before been performed.

First, we briefly provide an overview of the motivations and general philosophy used to create the proposed mine detection algorithm. The overall objective is to create a robust algorithm that will reliably detect underwater objects that pose a threat. However, the algorithm must also be fast for execution on an AUV with limited processing capabilities. Since no human intervention is allowed, the algorithm must be flexible enough to sense and adapt to changing environmental conditions from the data collected *in situ*. At the same time, the proposed method should directly address and overcome the limitations that plague existing detection algorithms. To satisfy these requirements, extensive domain-specific knowledge we possess about the problem is exploited. Emphasis is also placed on tailoring the algorithm, where possible, to the fundamental underlying physics and geometry of the application. The result is a hybrid of rigor and flexibility. Additionally, the cascaded algorithm architecture is designed to minimize computational costs by operating on progressively smaller portions of the image at each of three major stages: shadow detection, ripple detection, and echo detection.

Sections II-B–II-E will describe each step of the detection algorithm with enough detail to allow other researchers to replicate the method. And because the approach lends itself beautifully to visualization, the algorithm description will be supplemented by figures illustrating the effects of each step on one particularly challenging sonar image, shown in Fig. 1.

<sup>1</sup>There is an unavoidable delay due to the fundamental nature of SAS processing, in which multiple (future) pings are utilized to construct the synthetic array.

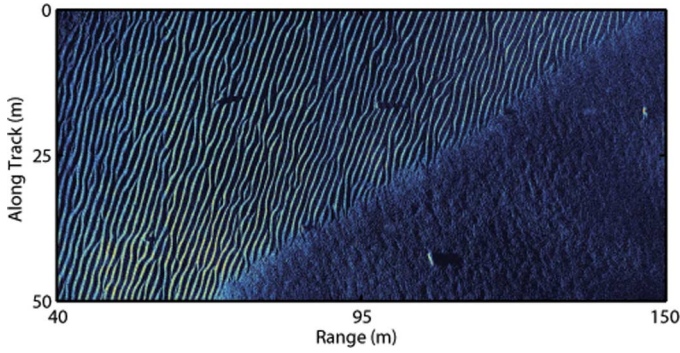


Fig. 1. SAS image that will be used as a running example for explaining each step of the detection algorithm. The left portion of the image is characterized by sand ripples, while the right portion is flat sandy seabed. Also present in the image are seven mine-like targets, the locations of which will be revealed later. The image comprises  $2001 \times 7333$  pixels.

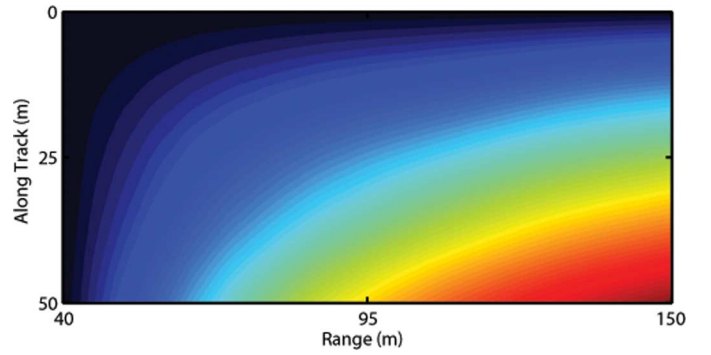


Fig. 2. Integral-image representation of the sonar image in Fig. 1.

### B. Shadow Detection

The first stage of the proposed detection algorithm establishes locations of shadows in the image. As an image preprocessing step, a simple median normalization is applied to the sonar image. (Specifically, let  $m_i^r$  be the median pixel value in the  $i$ th image row. Each pixel value in the  $i$ th row is then normalized by dividing by  $m_i^r$ , with this operation repeated for each  $i$ . Then, let  $m_j^c$  be the median pixel value in the  $j$ th image column. Each pixel value in the  $j$ th column is then normalized by dividing by  $m_j^c$ , with this operation repeated for each  $j$ .) The median-normalized sonar image is then converted to an integral-image representation. Subsequently, the integral image is used to quickly estimate a background map and a shadow map, the comparison of which provides locations of shadows (consistent with mines) that are examined further in the second stage of the algorithm.

1) *Integral Image*: An integral image [19] is an image representation that allows for very fast computation of rectangular, Haar-like features at any scale or location in constant time (since the computation does not depend on the size of the input). In subsequent stages of the algorithm, such rectangular features will be invaluable for assessing certain distinguishing characteristics of objects, such as echo and shadow levels. The use of an integral-image representation is also a key that makes *streaming* detection possible. (When dealing with the integral image calculations, care should be taken to ensure that sufficient numerical precision is used to avoid unintended truncation errors.)

Thus, rather than operating on the pixel-based system of the sonar image, we immediately transform to an integral-image system. Starting from an original sonar image  $A$ , the corresponding integral image  $I$  is constructed as follows. (It should be noted that, in the following, the variable  $x$  runs in the vertical direction and  $y$  runs in the horizontal direction.) The value at a location  $(x, y)$  in the integral image corresponds to the sum of the pixels above and to the left of  $(x, y)$ , inclusive, in the original image  $A$ . That is

$$I(x, y) = \sum_{x' \leq x, y' \leq y} A(x', y'). \quad (1)$$

The integral image is quickly generated using the recursive relation

$$I(x, y) = I(x - 1, y) + z(x, y) \quad (2)$$

where  $z(x, y)$  is the cumulative sum of pixels in a row of the original image

$$z(x, y) = z(x, y - 1) + A(x, y). \quad (3)$$

The integral-image representation corresponding to the sonar image in Fig. 1 is shown in Fig. 2.

2) *Background Estimation*: The first use of the integral image  $I$  is in the estimation of the sonar-image background map  $B$ . The purpose of the background map is to establish the reverberation level of the seabed to subsequently determine locations of shadows in the image.

The reverberation level is a strong function of the seabed composition. For instance, the reverberation level of a soft clay seabed will be lower than that of a seabed of hard-packed sand. In fact, natural seabed variations can cause the reverberation level to vary substantially at a given site or even within the same image.

Thus, in this work, a robust estimation of the background is obtained by using the *local* characteristics of the seabed in the measured data, as was also done in [20] and [21]. The inherent flexibility afforded by tailoring the estimation to the data itself improves the method's robustness for successfully dealing with different environments. The adaptive estimation also eliminates the possibility of training data mismatch [22], [23], since no training data are used.

The background estimation is calculated using a split-window template [24] composed of three equally sized (stacked) horizontal bands (i.e., rectangles) of which the top and bottom bands contribute to the background value, while the middle band does not. Each band is of the same size,  $b_x = 2.5$  m in the along-track direction and  $b_y = 5.0$  m in the range direction. These sizes are selected such that an object of interest (i.e., a mine) would be fully contained within the middle band, and therefore, not contribute to the background estimate. (For a different application concerned with other types of objects, these parameters would be adjusted accordingly.)

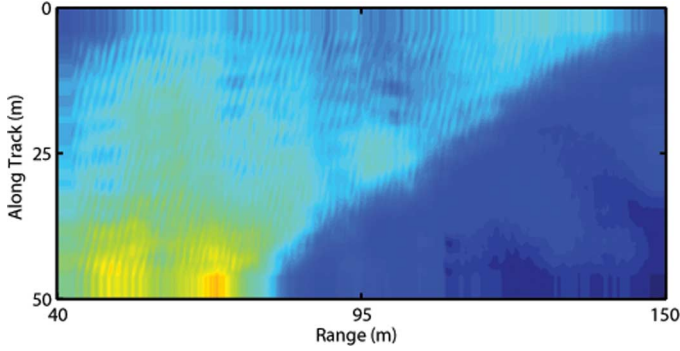


Fig. 3. Background map associated with Fig. 1.

The background score at location  $(x, y)$ ,  $B(x, y)$ , is then the mean pixel value in the two rectangles with the template centered around  $(x, y)$ . Thanks to the integral-image representation, the calculation of the two rectangles that contribute to the background value at a given location can be computed quickly, with a total of only eight array references from the integral image.

Specifically, the background score at  $(x, y)$  is calculated as

$$B(x, y) = a_b^{-1} \left[ I\left(x - 3\frac{\delta x}{2}, y - \frac{\delta y}{2}\right) - I\left(x - 3\frac{\delta x}{2}, y + \frac{\delta y}{2}\right) - I\left(x - \frac{\delta x}{2}, y - \frac{\delta y}{2}\right) + I\left(x - \frac{\delta x}{2}, y + \frac{\delta y}{2}\right) + I\left(x + \frac{\delta x}{2}, y - \frac{\delta y}{2}\right) - I\left(x + \frac{\delta x}{2}, y + \frac{\delta y}{2}\right) - I\left(x + 3\frac{\delta x}{2}, y - \frac{\delta y}{2}\right) + I\left(x + 3\frac{\delta x}{2}, y + \frac{\delta y}{2}\right) \right] \quad (4)$$

where  $\delta x$  and  $\delta y$  are the numbers of pixels that correspond to  $b_x$  and  $b_y$ , respectively, and  $a_b$  is the total number of pixels involved in the sums, so that the result is the mean pixel value from the two rectangular areas.

The region immediately surrounding a location (i.e., the “middle band”) is not used in the background calculation because those pixels, if an object were present, would be related to the object and bias the background estimate. Background *normalization* schemes [25] similarly attempt to exclude signal energy in calculations. The resulting background map from using the integral image in Fig. 2 is shown in Fig. 3, where it can be seen that the background level indeed varies across the image.

3) *Shadow Estimation*: Next, the integral image is used again to construct a shadow map. The purpose of the shadow map is to aid in the determination of the locations of shadows in the image that could have been produced by objects of interest.

Recall that the objective of the detection algorithm is to detect mines on the seafloor. The extensive domain-specific knowledge we possess—namely, the size and shape of likely mines—allows us to estimate the minimum height above the seafloor that such an object would extend. But the unique geometry of the problem, involving the AUV and sonar and objects on the seafloor, provides considerably more information

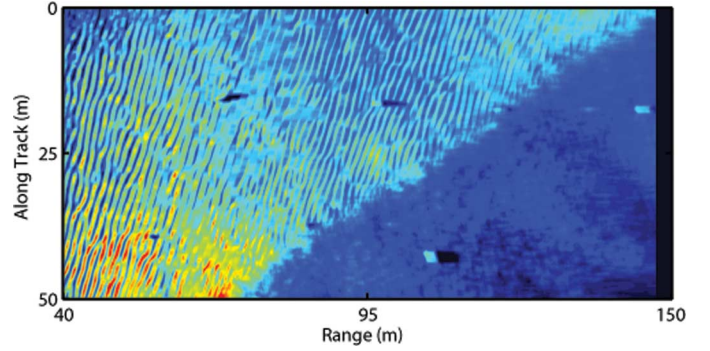


Fig. 4. Shadow map associated with Fig. 1.

that can be exploited in the detection algorithm. Any object that rises above the (assumed flat) seafloor will cast a shadow. Moreover, the length of the shadow that is cast will be a function of the relative positions of the object and the sonar.

For an AUV operating at an altitude  $a$  above the seafloor, and for a mine of height  $h$  sitting on the seafloor at a range  $r$  away from the AUV, simple geometry dictates that the length of the shadow that will be cast by the object will be

$$s_y = \frac{hr}{a - h}. \quad (5)$$

The value of  $h$  is set to the minimum height of an object that we are interested in detecting, drawn from a library of relevant mines (here,  $h = 0.25$  m is used);  $a$  is measured directly onboard the AUV.

To respect the inviolable geometry of the problem, the size of the rectangular template used for the shadow calculation is made range dependent, effectively refining the search for simple shadows to a search for shadows of an *appropriate length*. This choice instills an additional layer of rigor into the method. The size of the rectangle used in the calculation is  $s_x = 0.5$  m in the along-track direction, but  $s_y$  from (5) in the range direction.

The shadow score at  $(x, y)$  is taken to be the mean pixel value over the rectangle [centered around  $(x, y)$ ], again calculated quickly thanks to the integral-image representation

$$S(x, y) = a_s^{-1} \left[ I\left(x - \frac{\delta x}{2}, y\right) - I\left(x - \frac{\delta x}{2}, y + \delta y\right) - I\left(x + \frac{\delta x}{2}, y\right) + I\left(x + \frac{\delta x}{2}, y + \delta y\right) \right] \quad (6)$$

where  $\delta x$  and  $\delta y$  are the numbers of pixels that correspond to  $s_x$  and  $s_y$ , respectively, and  $a_s$  is the total number of pixels involved in the sum, so that the result is the mean pixel value from the rectangular area. The resulting shadow map from using the integral image in Fig. 2 is shown in Fig. 4.

It should be noted that because of the integral-image representation, the range-dependent nature of the template causes no additional computational complexity. In contrast, existing detection algorithms that employ matched-filtering-type methods that rely on discrete Fourier transforms cannot enjoy this range-dependent flexibility. Instead, those approaches will be inherently mismatched (since the geometry of the problem is not respected) and inferior; to wit, either a single average range (which biases against detecting targets at certain ranges) is used

or the whole process is repeated multiple times for different expected shadow sizes, which increases the computational time.

4) *Addressing Poor Image Quality*: A sonar image with good image quality is one in which objects and seabed features appear more or less as they would when observed under the ideal operating conditions of the sonar system. It is commonly assumed that the full extent of all sonar images are of good quality, but that is often not the case with real sonar data collected at sea [26].

In particular, image quality often degrades significantly at long range because of spreading and absorption (i.e., attenuation), which leads to weakened signal strength, or multipath effects [27]. Additionally, with SAS, the motion correction requirements for successful image formation are more stringent at long range [28]. As a result, the image quality of data at the longest ranges of the sonar system is often inadequate.

In this poor-image-quality regime, shadow purity becomes diminished (i.e., shadows “fill in”), making shadows an unreliable clue for object detection. Therefore, in the proposed algorithm, all regions in the image where the image quality is poor are automatically retained to be examined further in the subsequent stages. What constitutes poor image quality is determined as follows.

SAS works by coherently summing received sonar signals, or ping returns, of overlapping elements in the array. The cross correlation between two received discrete-time signals  $s_i$  and  $s_j$  is

$$\kappa_{ij}[n] = \sum_{m=-\infty}^{\infty} s_i[m] s_j^*[n+m] \quad (7)$$

where  $n$  is a time shift (and in practice the summation is computed for finite bounds dictated by the signals’ lengths). The normalized peak correlation between successive pings, termed the *coherence* [29], is then

$$\rho = \max_n \left| \frac{\kappa_{ij}[n]}{\sqrt{\kappa_{ii}[0] \kappa_{jj}[0]}} \right|. \quad (8)$$

In this work, the mean coherence value of consecutive ping returns, as a function of range, is employed as a surrogate measure of the image quality. This quantity is computed within the displaced phase-center antenna (DCPA) algorithm [28] of the SAS processing and is directly related to the signal-to-noise ratio (SNR) [30].

Specifically, the coherence at range  $r$  is estimated as follows. The portion of the signal return that corresponds to a window of ranges  $[r - \Delta r, r + \Delta r]$  about range-center  $r$  can be determined when an estimate of the sound speed is possessed. These signal portions are used in the coherence calculation above, resulting in  $\rho_i(r)$  for the  $i$ th of  $P$  pairs of consecutive pings that contribute to the image. The mean of the coherence values  $\rho(r) = (1/P) \sum_{i=1}^P \rho_i(r)$  is then computed, with this averaging performed across the ping pairs to improve robustness at the expense of along-track granularity. If the mean coherence  $\rho(r)$  is below a set threshold, the image indices  $(x, y)$  associated with ranges  $[r - \Delta r, r + \Delta r]$  are deemed to be of poor image quality. In each image, this procedure is performed for a discrete set of range centers  $r$ , with  $\Delta r = 5$  m typically.

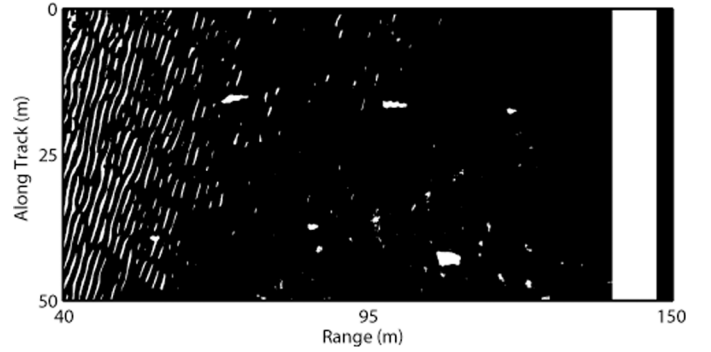


Fig. 5. Result of comparing the background map of Fig. 3 and the shadow map of Fig. 4 is a binary map indicating shadow regions from potential alarms. That result is then combined with the binary image-quality map to produce this overall map showing regions of interest to be examined further.

Based on our extensive experience visually inspecting sonar images, we deem an image at ranges where  $\rho(r)$  is below the threshold  $\tau_\rho = 2/3$  (corresponding to a 3-dB SNR) to be of poor image quality containing unreliable shadow information. (A similarly *ad hoc* threshold of  $\tau_\rho = 3/4$  is adopted in [29] for declaring image quality poor.) Thus, with a threshold selected, a binary image-quality map can be obtained readily.

5) *Region of Interest Determination (Stage I)*: After a background map  $B$  and a shadow map  $S$  are constructed, the regions of the image that might contain targets of interest are determined. Any pixel for which the shadow map value is sufficiently lower than the corresponding background map value is declared to be a region of interest that will receive further investigation. Specifically, if  $S(x, y) < \beta B(x, y)$ , then the pixel  $(x, y)$  is considered to be part of a shadow and, therefore, a region of interest that will receive further investigation. The scaling factor  $\beta$  makes the requirement for shadows more stringent (here,  $\beta = 0.5$ ). Essentially, this *defines* that a pixel is part of a shadow region when the value is a certain amount less than the average value of the surrounding background.

The binary result of the background and shadow map comparison, combined with the binary image-quality map from Section II-B4, is shown in Fig. 5, where white pixels correspond to regions of potential alarms that will be investigated further. The vertical stripe of white at longer range (which is where the image quality was poor) is the contribution from the image-quality map.

By employing the background map in the search for shadows, local contextual information is being exploited. Algorithm robustness is also enhanced by focusing on an intrinsic aspect of the problem: objects of interest must rise a certain height above the seafloor, which means a shadow of a particular length will be cast. (The algorithm is not intended for completely buried mines.)

The background and shadow map comparison is the first data-reduction stage of the detection cascade. It achieves a large-scale reduction in pixels that must be examined further (in practice, more than 90% of the SAS image pixels are usually removed in this step), thereby greatly reducing computation loads of all subsequent stages. (Recall that the original SAS image comprised over 14 million pixels.) Since naturally occurring highlights (not caused by objects extending above the seabed)



in sonar images are common, instead searching first for locations of high signal strength would not enable as significant a data reduction.

### C. Ripple Detection

The second stage of the proposed detection algorithm detects ripples in the sonar image. A feature extraction step maps each shadow to a point in a new space in which ripples, characterized by a high density of elongated shadows that are oriented similarly, can be efficiently detected. Locations that do not belong to ripples are examined further in the third stage of the algorithm.

1) *Motivation*: Sand ripples are one of the most confounding factors and frequent sources of false alarms for underwater mine detection because they produce highlight-shadow patterns—the sizes of which will depend on the amplitude, period, and range of the ripples—characteristic of mine-like objects. Additionally, when mines lie in ripple fields, the echoes of the mines and ripples can blend together, as can their respective shadows, making detection challenging.

Since the overall objective here is to detect mines, the ripple detection algorithm we propose is more accurately called a *nuisance* ripple detection algorithm. Rather than attempting to detect all sand ripples in an image, the algorithm instead attempts to detect only those ripples that affect our ability to detect mines, i.e., nuisance ripples.<sup>2</sup>

2) *Overview*: The principled ripple detection algorithm developed here is tailored both to the fundamental characteristics that *define* sand ripples and the underlying physics that creates them. Sand ripples are formed when currents or waves flow in a particular direction [31]–[33]. Because of this underlying mechanism by which they are created, ripples tend to exist in large *fields*, rather than as individual instantiations. But a ripple field, by definition, must contain a certain high density of ripples.

Although the ripples will be characterized by a dominant orientation (dictated by the direction of flow), considerable variability will also exist. That is, ripples will not manifest at a single fixed orientation. Rather, they will span a range of orientations about some dominant one. Moreover, the amplitudes and periods of the ripples can also vary. Therefore, to be able to detect such variability, a method that is inherently flexible is required. Additionally, the relative geometry between the sonar and the mound of sand that constitutes a ripple means that the shadows that are cast by the ripples will be elongated along the direction of the ripple crest (or trough) [34].

These three fundamental physical characteristics of ripples, which can be observed in Fig. 1, will be exploited in the ensuing stages of the ripple detection algorithm. Specifically, the proposed ripple detection algorithm searches for a high density of elongated shadows that are oriented similarly. To perform this search efficiently, each discrete shadow region obtained from the background-shadow map comparison (cf., Fig. 5) is mapped into a new feature space in which it is easy to detect such characteristics. A novel, unconventional use of the integral-image formulation for purposes of density estimation is then employed.

<sup>2</sup>To instead detect *all* ripples, such as for environmental assessment purposes, the template used in the shadow map determination should be made range independent.

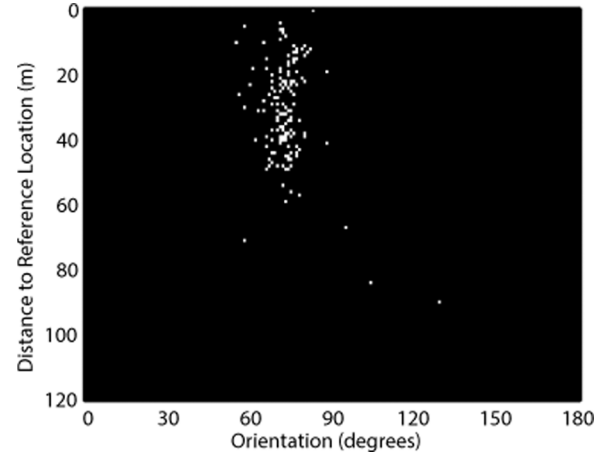


Fig. 6. Each shadow region from Fig. 5 is mapped to a point in location–orientation space after a location feature and an orientation feature are extracted.

3) *Feature Extraction*: To begin, the centroid and orientation is calculated for each discrete shadow (i.e., each connected region, or “blob”) in the binary shadow map. (No morphological operators are used to alter the shadow map.) The orientation (in degrees)  $\phi$  of a shadow is obtained by fitting an ellipse to the shadow area, and then correcting for unequal pixel sizes. The centroid is converted to a distance feature (in meters)  $\chi$  that is simply the distance of the centroid to a fixed reference location in the mission (for the case of a single image here, the “first” pixel of the image: the pixel at the nearest range at the initial along-track position). This feature extraction process effectively maps each shadow to a specific location  $(\chi, \phi)$  in a new feature space in  $\mathbb{R}^2$ . (To treat this space as an image later, these features are then discretized to  $1^\circ$  in the orientation–dimension and 1 m in the location–dimension.)

Before proceeding, though, certain shadows are removed from further consideration. If  $\phi < d_\phi/2$  or  $\phi > 180^\circ - d_\phi/2$ , an indication that the shadow is nearly horizontal, the shadow is ignored. (In this work,  $d_\phi = 30^\circ$ .) This choice is made because it is known *a priori*, from the geometry of the sonar and seabed, that the majority of shadows at those prohibited near-horizontal orientations are likely caused by mine-like objects (as opposed to ripples). Additionally, this choice eliminates any inconsistency that would arise from not respecting the “wraparound” nature of orientations (from  $180^\circ$  to  $0^\circ$ ) in the upcoming integral-image density estimation. The result of this feature extraction process that transforms the shadows in sonar-image space in Fig. 5 to location–orientation space is shown in Fig. 6.

4) *Density Estimation*: By treating this new feature space as a 2-D “image”  $A'$ , the corresponding integral image  $I'$  can be computed readily. The integral-image formulation can then be exploited to quickly perform density estimation in this space. Specifically, the mean number of discrete shadows in a rectangular area about a given location is used as the shadow density estimate.

The size of the rectangle used in the calculation is  $d_\chi = 25$  m in the location–dimension and  $d_\phi = 30^\circ$  in the orientation–dimension. This size is chosen because it respects the physical proximity and orientation variability of the underlying mechanisms that help define ripple fields. The density estimate at

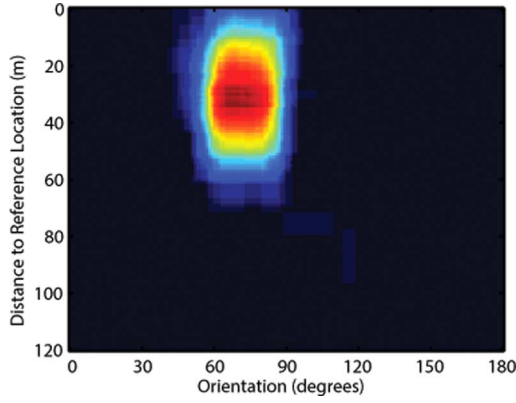


Fig. 7. Shadow density in location–orientation space from Fig. 6.

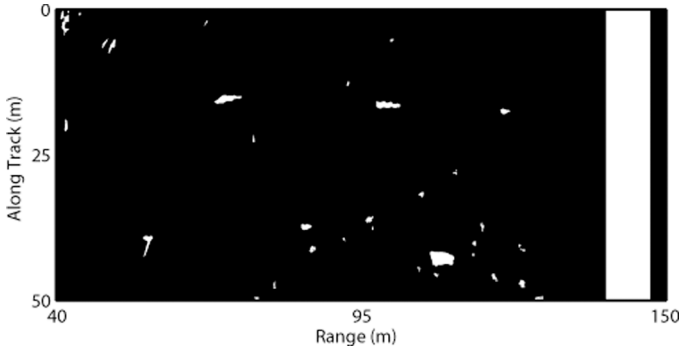


Fig. 8. Potential alarm regions that remain after the ripple detector removes regions by exploiting Fig. 7.

$(\chi, \phi)$ ,  $D(\chi, \phi)$ , is taken to be the mean pixel value (though irrespective of orientation) over the rectangle centered around  $(\chi, \phi)$ , calculated quickly as

$$D(\chi, \phi) = d_{\chi}^{-1} \left[ I' \left( \chi - \frac{d_{\chi}}{2}, \phi - \frac{d_{\phi}}{2} \right) - I' \left( \chi - \frac{d_{\chi}}{2}, \phi + \frac{d_{\phi}}{2} \right) - I' \left( \chi + \frac{d_{\chi}}{2}, \phi - \frac{d_{\phi}}{2} \right) + I' \left( \chi + \frac{d_{\chi}}{2}, \phi + \frac{d_{\phi}}{2} \right) \right]. \quad (9)$$

Although a span of orientations are included in the density estimate, the mean is calculated over only the location–dimension because we do not wish to distinguish among ripples at similar orientations (since minor orientation variations naturally exist in a general ripple field). The resulting shadow density estimate, calculated from Fig. 6, is shown in Fig. 7.

5) *Classification*: If the maximum density is above a predefined threshold  $\tau_d$ , sand ripples are declared to be present in the original sonar image. The threshold  $\tau_d$  effectively defines the minimum density of shadows, oriented similarly, needed to constitute a ripple field. (In this work,  $\tau_d = 0.5$ , which dictates that at least 12 shadows at similar orientations within a given distance of each other are needed to permit a declaration of a ripple field.)

The principal (center) orientation of the ripple field is estimated to be the orientation at which the density achieves a maximum. Additionally, the ripple field is deemed to span a set of orientations, with the minimum and maximum orientations of that span set to the orientations at which the density drops to

half the maximum density. That is, the span is defined adaptively based on the evidence supported by the data itself.

For the image related to Fig. 7, the principal ripple orientation was declared to be  $\theta_r = 65^\circ$ , spanning the orientations  $\Theta_r = [57^\circ, 87^\circ]$ . Comparison with the original sonar image in Fig. 1 indicates that the ripple orientations were estimated accurately.

6) *Region of Interest Determination (Stage II)*: If sand ripples have been detected, those potential alarms for which the orientation is within the estimated span of ripple orientations  $\Theta_r$  are removed from further consideration. The result of the nuisance ripple detection stage is shown in Fig. 8, from which it can be observed that the ripples have largely been removed. This test is the second data-reduction stage of the detection cascade, and further reduces the number of pixels that must be examined subsequently.

#### D. Echo Detection

The third stage of the proposed detection algorithm estimates the echo (highlight) strength of the remaining alarms. A range-dependent correction term to counteract grazing angle effects is also applied before determining the final list of discrete mine-like alarms.

1) *Echo Estimation*: Next, an echo score is calculated for the remaining potential alarm regions by again exploiting the integral image. Specifically, the mean pixel value in a rectangular area about a given location is used as the echo score.

The size of the rectangle used in the calculation is  $e_x = 0.75$  m in the along-track direction and  $e_y = 0.5$  m in the range direction. This size is chosen because it is related to the smallest object of interest we wish to detect. The echo score at  $(x, y)$  is taken to be the mean pixel value over the rectangle, calculated as

$$E(x, y) = a_e^{-1} \times \left[ I \left( x - \frac{\delta x}{2}, y - \frac{\delta y}{2} \right) - I \left( x - \frac{\delta x}{2}, y + \frac{\delta y}{2} \right) - I \left( x + \frac{\delta x}{2}, y - \frac{\delta y}{2} \right) + I \left( x + \frac{\delta x}{2}, y + \frac{\delta y}{2} \right) \right] \quad (10)$$

where  $\delta x$  and  $\delta y$  are the numbers of pixels that correspond to  $e_x$  and  $e_y$ , respectively, and  $a_e$  is the total number of pixels involved in the sum, so that the result is the mean pixel value from the rectangular area.

However, a minor realignment must be effected before the echo score calculations because the potential alarm regions are located at the positions of the *shadows* cast by the objects. Since the echo scores that we desire should correspond to the *highlights* produced by the objects, which are necessarily at shorter range (i.e., closer to the sensor) than the shadows, the locations at which the echo score is computed are all shifted in range by  $3e_y/2$ . The resulting echo map from using the integral image in Fig. 2 is shown in Fig. 9, where it can be seen that the calculation is performed for only the relevant subset of pixels still under consideration.

2) *Side-Echo Estimation*: A “side-echo” estimate is used to assess whether an echo is sufficiently stronger than the echo of its local surroundings. The side-echo estimation is calculated

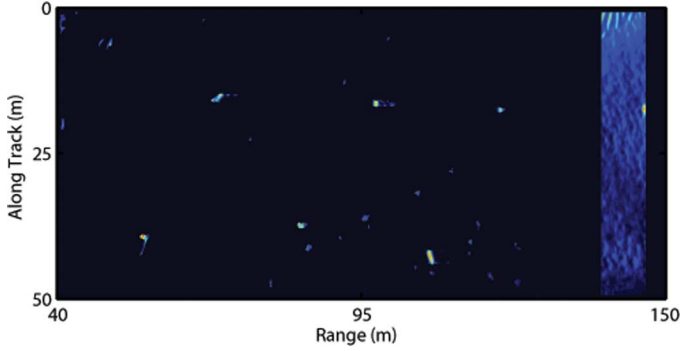


Fig. 9. Echo map associated with Fig. 1 over the subset of pixels specified in Fig. 8.

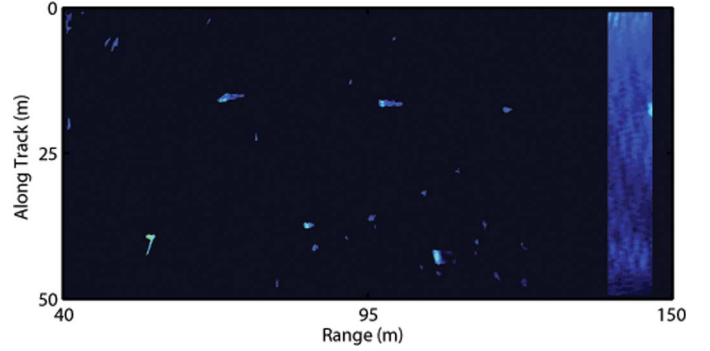


Fig. 10. Side-echo map associated with Fig. 1 over the subset of pixels specified in Fig. 8.

using a split-window template similar to the one used for the background calculation, except here the middle band is twice the height of each of the other two bands. As before, only the top and bottom bands contribute to the side-echo value, while the middle band does not. The top and bottom bands, or rectangles, are of the same size,  $f_x = e_x/2 = 0.375$  m in the along-track direction and  $f_y = 2e_y = 1.0$  m in the range direction. Since the middle band is  $e_x = 0.75$  m in the along-track direction, the region used in the echo calculation from Section II-D1 is fully contained within this middle band.

The side-echo score at location  $(x, y)$ ,  $F(x, y)$ , is then the mean pixel value in the two rectangles with the template centered around  $(x, y)$ . Thanks to the integral-image representation, the calculation of the two rectangles that contribute to the side-echo score at a given location can be computed quickly, with a total of only eight array references from the integral image.

Specifically, the side-echo score at  $(x, y)$  is calculated as

$$\begin{aligned}
 F(x, y) &= a_f^{-1} \left[ I \left( x - \frac{\delta x}{2} - \Delta x, y - \frac{\Delta y}{2} \right) \right. \\
 &\quad - I \left( x - \frac{\delta x}{2} - \Delta x, y + \frac{\Delta y}{2} \right) \\
 &\quad - I \left( x - \frac{\delta x}{2}, y - \frac{\Delta y}{2} \right) + I \left( x - \frac{\delta x}{2}, y + \frac{\Delta y}{2} \right) \\
 &\quad + I \left( x + \frac{\delta x}{2}, y - \frac{\Delta y}{2} \right) - I \left( x + \frac{\delta x}{2}, y + \frac{\Delta y}{2} \right) \\
 &\quad - I \left( x + \frac{\delta x}{2} + \Delta x, y - \frac{\Delta y}{2} \right) \\
 &\quad \left. + I \left( x + \frac{\delta x}{2} + \Delta x, y + \frac{\Delta y}{2} \right) \right] \quad (11)
 \end{aligned}$$

where  $\delta x$  is the number of pixels that corresponds to  $e_x$ ;  $\Delta x$  and  $\Delta y$  are the numbers of pixels that correspond to  $f_x$  and  $f_y$ , respectively; and  $a_f$  is the total number of pixels involved in the sums, so that the result is the mean pixel value from the two rectangular areas.

The region immediately surrounding a location (i.e., the “middle band”) is not used in the side-echo calculation because those pixels contribute to the echo score with which we make the comparison. As with the echo score calculation, the locations at which the side-echo score is computed are

all shifted closer in range by  $3e_y/2$ , where  $e_y$  was defined in Section II-D1. The resulting side-echo map from using the integral image in Fig. 2 is shown in Fig. 10, where again the calculation is performed for only the relevant subset of pixels still under consideration.

3) *Echo-Loss Correction*: The backscattering strength of the seabed tends to decrease at lower grazing angles [35]. Since peak target strengths tend to be relatively independent of grazing angle, and grazing angles decrease as range increases, the signal-to-reverberation ratio for target echoes tends to decrease at shorter ranges. Consequently, the median normalization that is performed on the original sonar image as a preprocessing step will have the effect of reducing returns from shorter ranges more substantially than those from longer ranges. In turn, this means that the normalization step would systematically cause the echo level of a target at short range to be lower than the echo level of the same target if it had been at longer range.

To ameliorate this undesirable effect, which would make targets at shorter ranges more difficult to detect, we apply a generic range-dependent correction term that is intended to be suitable for various seabed types. The *ad hoc* echo correction that we add to the echo map (and side-echo map) at range  $r$  is

$$\epsilon(r) = 20 \log_{10} \left( \frac{r_0}{r} \right) \quad (12)$$

where  $r_0$  is a reference range set to the nominal maximum range of the sonar (here,  $r_0 = 150$  m). This correction is applied only for potential alarm regions, which again saves computation. A more principled physics-based formulation for this correction is an item for future work.

4) *Echo Comparison*: After an echo map  $E$  and a side-echo map  $F$  are constructed and the echo-correction has been applied, the potential alarm regions of the image are revised as follows. Any pixel for which the echo map value is sufficiently higher than the corresponding side-echo map value is declared to be a region of interest that will receive further investigation. Specifically, if  $E(x, y) > \gamma F(x, y)$ , then the pixel  $(x, y)$  is considered to be part of an object echo and, therefore, a region of interest that will receive further investigation. The scaling factor  $\gamma$  makes the requirement for echoes more stringent (here,  $\gamma = 2.5$ ). Essentially, this defines that an object echo is when the pixel value is a certain amount more than the average value of the surrounding area, or seabed echo. The binary map resulting



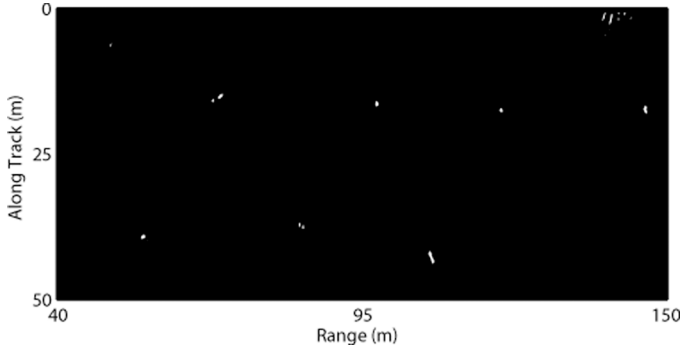


Fig. 11. Potential alarm regions that remain after comparing the echo-corrected versions of the echo map of Fig. 9 and the side-echo map of Fig. 10.

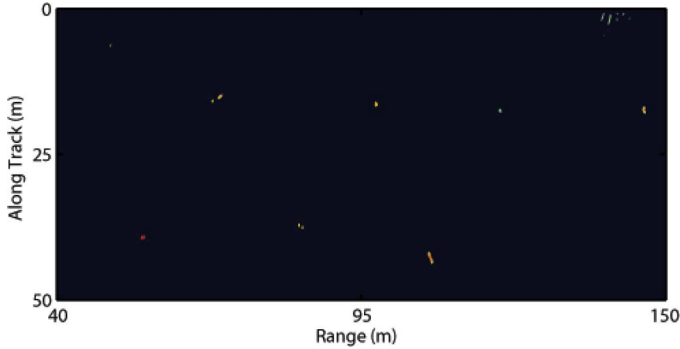


Fig. 12. Corrected echo map associated with Fig. 1 over the subset of pixels specified in Fig. 11.

from the echo map and side-echo map comparison is shown in Fig. 11, where white pixels correspond to regions of potential alarms that will be investigated further.

This comparison test is another data-reduction stage of the detection cascade. It removes from consideration those areas that had indications of shadow but not an echo significantly stronger than the seabed. The corresponding (corrected) echo map that results is shown in Fig. 12.

5) *Region of Interest Determination (Stage III)*: The echo-correction step ensures that the echo score of a given target should be approximately range independent. In turn, this permits the use of a single detection threshold  $\tau_e$  for all ranges. Since the echo scores are directly related to the highlight intensity of the objects, this threshold can be set to enable the detection of objects with a given highlight level. The final stage of the detection cascade removes those regions associated with an echo score below the desired threshold (here,  $\tau_e = 8$ ). The map of potential alarm regions that remain after this echo test is shown in Fig. 13.

6) *Density Filtering*: The next step of the detection algorithm is to convert the areas of potential alarms to a list of *discrete* alarms. To do this, the location of the maximum echo score within each potential alarm region is treated as the alarm location. Then, contextual information is exploited in an optional filtering step to reject alarms that collectively, over an area of a prescribed size, constitute an unrealistically high density of contacts. In this work, if  $n_d \geq 4$  alarms are contained within an area of  $a_d = 15 \text{ m} \times 15 \text{ m}$ , those alarms are eliminated from consideration. The alarms that remain after this density filtering step constitute the final list of detections. (For the example image, the density filtering step had no effect.)

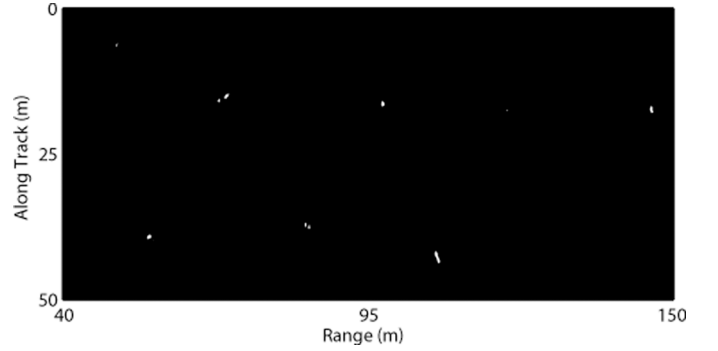


Fig. 13. Potential alarm regions that remain after regions in Fig. 12 with weak echoes have been removed by the echo test.

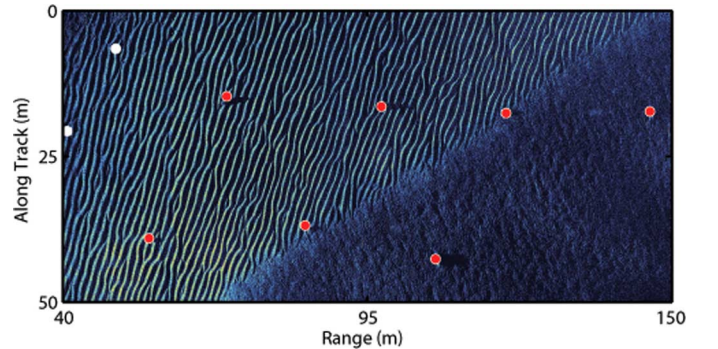


Fig. 14. Final detection map obtained from converting alarm areas in Fig. 13 into discrete alarms. The solid white circles indicate false alarms, while red/hollow-white circles mark correct target detections.

To present the final detection result visually for our example image, we plot the alarm locations overlaid on the original sonar image in Fig. 14. It can be seen that all seven targets were successfully detected but two false alarms were recorded. For purposes of comparison, a matched-filtering-type detection method [8] applied to this image generated 112 false alarms—most due to the sand ripples—and also failed to detect three targets (two in the ripple field and one at long range where the image quality is poor).

#### E. Final Detection Score

The last step of the detection algorithm is to assign a final confidence score for each detected alarm. (This procedure is intended simply to *order* the list of alarms, not to reject additional alarms as would a follow-on classification stage.) An earlier, preliminary version of the detection algorithm [18] employed the maximum echo score of an alarm (obtained in Section II-D6) as the final detection score, but here a new feature is proposed. The final detection score of an alarm is formed as the average of two new features, which we call the highlight concentration and the shadow concentration. Taken together, these two basic features accurately express the degree to which an alarm looks generically mine-like. Essentially, the highlight concentration feature measures the proportion of highlight pixels in an alarm's "mugshot" that is contained in an area in which a mine's highlight is expected to occur. The shadow concentration feature measures an analogous quantity for shadows.

For a detected object, let  $A^M$  denote a Gaussian-smoothed version of its  $5 \text{ m} \times 10 \text{ m}$  mugshot extracted from the SAS image scene  $A$ . Let  $A^H$  and  $A^S$  denote subimages extracted

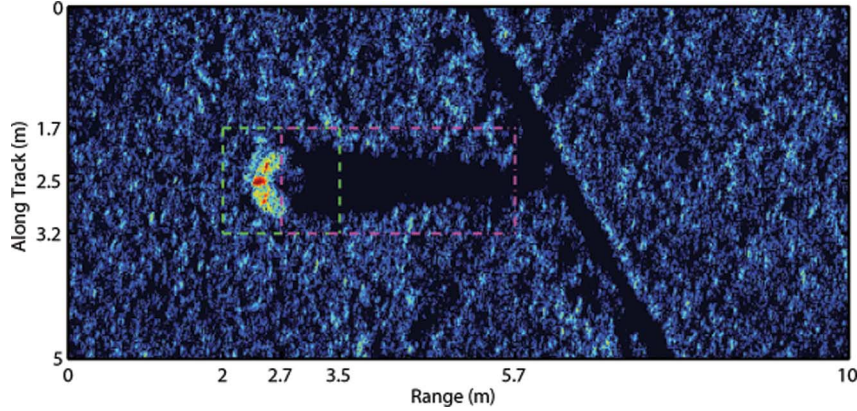


Fig. 15. “Mugshot” of an alarm with the expected highlight region and the expected shadow region—used for feature calculations—marked in green/dashes and magenta/dash-dots, respectively. (N.B. The range values are offset by 100 m in the figure for clarity.)

from  $A^M$  that cover what we refer to as an expected highlight region and an expected shadow region, respectively; the boundaries of these particular regions are shown by way of example in Fig. 15. It should be noted that the sizes and locations of these regions are fixed, identical for all mugshots. Here,  $A^H$  is  $1.5 \text{ m} \times 1.5 \text{ m}$  and  $A^S$  is  $1.5 \text{ m} \times 3 \text{ m}$ , with these sizes dictated by the sizes of objects of interest.

Let  $|H|$ ,  $|S|$ , and  $|M|$  denote the numbers of pixels in the expected highlight region, the expected shadow region, and the entire mugshot, respectively. The highlight concentration feature and shadow concentration feature are then calculated as

$$h_c = \frac{\sum_{i=1}^{|H|} \mathbb{I}(A_i^H \geq \tau_h)}{\sum_{j=1}^{|M|} \mathbb{I}(A_j^M \geq \tau_h)} \quad (13)$$

$$s_c = \frac{\sum_{i=1}^{|S|} \mathbb{I}(A_i^S \leq \tau_s)}{\sum_{j=1}^{|M|} \mathbb{I}(A_j^M \leq \tau_s)} \quad (14)$$

respectively, where  $\tau_h$  and  $\tau_s$  are fixed thresholds,  $\mathbb{I}$  is an indicator function that is unity when its argument is true (and zero otherwise), and  $A_i$  is the value of the  $i$ th pixel in the image specified by the superscript. The highlight threshold is set to  $\tau_h = 10$ . The shadow threshold  $\tau_s$  is set to the value corresponding to the  $q$ th quantile of  $A^M$ 's pixels, where  $q = |S|/4|M|$ . (Adapting the shadow threshold, from mugshot to mugshot, accounts for variation in shadow purity.) It can be readily seen that  $0 \leq h_c \leq 1$  and  $0 \leq s_c \leq 1$ .

The elegantly simple feature calculations in (13) and (14) mean that no computationally intensive segmentation algorithm [36] is required to determine highlight and shadow regions associated with the object. The final detection score of an alarm is then taken to be the average of the two concentration features

$$f_c = \frac{(h_c + s_c)}{2}. \quad (15)$$

Thus, if the highlight pixels of a mugshot are concentrated in the expected highlight region, and the shadow pixels of a mugshot are concentrated in the expected shadow region, the final detection score will be high, indicative of a mine. (For the alarm

shown in Fig. 15,  $h_c = 0.80$  and  $s_c = 0.72$ , so  $f_c = 0.76$ , which is a relatively high score.) As will be shown later, the new feature is much better than the echo score at assigning higher confidences to mine-like alarms; this has important consequences for *in situ* contact reinspection with an AUV.

### III. DATA

#### A. Sea Experiments

The performance of the proposed detection algorithm is evaluated using real, measured SAS data collected at sea. All of the data used in this study were collected by the CMRE's SAS-equipped MUSCLE AUV. The center frequency of the SAS is 300 kHz, and the bandwidth is approximately 60 kHz. The system enables the formation of high-resolution sonar imagery with a theoretical across-track resolution of 1.5 cm and a theoretical along-track resolution of 2.5 cm. The present study spans 29 880 SAS images, one example of which was shown in Fig. 1.

Specifically, this study examines target-detection performance using data collected during six major sea experiments that were conducted by the Centre for Maritime Research & Experimentation (CMRE), NATO Science and Technology Organization (La Spezia, Italy) between 2008 and 2012; the experiments were Colossus 2, CATHARSIS 1, CATHARSIS 2, AMiCa, ARISE 1, and ARISE 2. Because the environments of the three principal areas in the Colossus 2 experiment were markedly different, the subsequent analysis treats these areas individually rather than as part of the same monolithic group. This choice allows us to better examine the effects of different environmental conditions on detection capability. Collection details about the data sets are given in Table I.

#### B. Targets

In each sea experiment, different groups of man-made targets were laid on the seabed. Surveys with the MUSCLE AUV were then performed over the target areas (as well as other areas). In total, the area of seabed spanned by the SAS imagery collected during the six sea experiments was approximately 160 km<sup>2</sup>. Detection performance is assessed for three surrogate target shapes: cylinders, truncated cones, and wedge-shape objects. Due to the surrogate nature of the targets, the performance results presented in Section IV should be considered as comparative within the confines of this paper, and should not

TABLE I  
DATA SET COLLECTION DETAILS

Sea Experiment - Area	Code	Location	Year	Days/Month	SAS Images
Colossus 2 - D	COL2D	Riga, Latvia	2008	15-19/4, 3/5	1956
Colossus 2 - B	COL2B	Liepaja, Latvia	2008	26/4-2/5	3338
Colossus 2 - C	COL2C	Liepaja, Latvia	2008	20-25/4	3241
CATHARSIS 1	CAT1	Palmaria, Italy	2009	11/3-6/4	2733
CATHARSIS 2	CAT2	Elba, Italy	2009	5-20/10	2647
AMiCa	AMI1	Tellaro, Italy	2010	14/5-11/6	2908
ARISE 1	ARI1	La Spezia, Italy	2011	2/5-1/6	8309
ARISE 2	ARI2	Elba, Italy	2012	15/10-2/11	4748

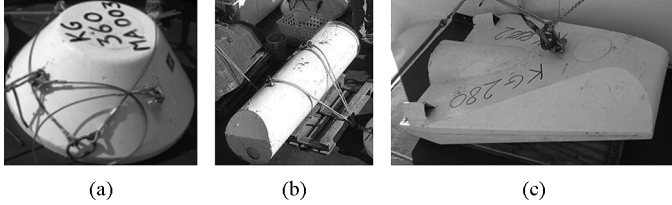


Fig. 16. Photographs of the target shapes considered in the study. (a) Truncated Cone. (b) Cylinder. (c) Wedge.

TABLE II  
DATA SET DETAILS RELEVANT FOR DETECTION

Sea Experiment - Area	Detection Opportunities	Seabed Area (km <sup>2</sup> )	Notable Environmental Characteristics
Colossus 2 - D	162	10.76	clay, flat sand
Colossus 2 - B	152	17.81	flat sand
Colossus 2 - C	141	18.34	sand ripples, flat sand, boulders
CATHARSIS 1	27	15.03	clay, flat sand
CATHARSIS 2	89	13.89	flat sand, boulders, posidonia
AMiCa	327	15.95	flat sand
ARISE 1	99	42.61	sand ripples, flat sand
ARISE 2	108	26.10	flat sand, boulders, posidonia
[All]	1105	160.49	–

be considered indicative of true mine hunting performance. Example photographs of the objects are shown in Fig. 16.

The total number of target detection opportunities (counting only these three principal target shapes) was 1105. Details about the detection opportunities in each data set, along with notable environmental characteristics relevant to the detection task, are given in Table II.<sup>3</sup>

### C. Example SAS Mugshots

To illustrate the variability of SAS target signatures that the detection algorithm is tasked to handle, example SAS-image mugshots of targets are shown in Fig. 17. In the figure, each column corresponds to one of the three target types (truncated cone, cylinder, wedge). The first row of images shows easy cases in which the targets appear more or less as their ideal forms; the examples shown were detected successfully. The second row of images shows cases that are more challenging for various reasons. In Fig. 17(d), the target shadow blends in

<sup>3</sup>The posidonia referred to in the table and throughout the text is the seagrass known as *Posidonia oceanica*, which is endemic to the Mediterranean Sea. It should be understood that a seabed described as posidonia is not referring to the seabed composition.

with the sand-ripple shadow. In Fig. 17(e), the image suffers from poor image quality so the weak target shadow has “filled in.” In Fig. 17(f), the target is on the edge of sand ripples and flat sand, with a relatively weak echo. Despite these complications, these targets were also detected successfully. The third row of images shows challenging cases in which the detection algorithm failed to detect the targets. In Fig. 17(g), the target is in posidonia and the echo is very weak. In Fig. 17(h), the target is well aligned with a highlight of a sand ripple and the target’s shadow is also weak. In Fig. 17(i), the target is on flat sand but the echo is very weak.

## IV. RESULTS

The term “probability of detection” is used here with the understanding that it means the probability of detecting a target *in a single view*. That is, each detection opportunity of a given target is treated independently. This point must be made explicit because often the same unique target is viewed multiple times (i.e., has multiple detection opportunities) in a given sea experiment. Thus, with the convention stated above, if a target was viewed in ten surveys but detected in only three of them, the probability of detection reported here would be  $p_d = 0.3$ . (In contrast, the probability the target would be detected *at least once* in ten opportunities, assuming independence, would be  $1 - \prod_{i=1}^{10} (1 - p_d) \approx 0.97$ .)

The proposed detection algorithm described in Section II was applied to all of the SAS images from each of the eight data sets noted in Section III. In addition to validating the proposed algorithm, the following large-scale performance assessment also provides significant insight regarding the quantities that influence detection capability in SAS imagery. To this end, the performance of the detection algorithm is first examined as a function of target type, seabed environment, and geographical site. Subsequently, the effects of image quality and the final detection score on performance are studied. Detection performance as a function of target aspect and range is also considered. Then, the impact of sand-ripple orientation is discussed. Unless otherwise noted, all results correspond to those obtained with the standard  $\tau_p = 2/3$  setting for the image-quality threshold within the detection algorithm. No attempt has been made to optimize the settings of the algorithm’s various parameters.

### A. Target Type, Seabed, and Site Dependence

The performance of the proposed detection algorithm as a function of target type and seabed environment is summarized in Table III. The results clearly indicate that the dominant factor in determining whether a target will be detected is the environmental characteristics of the surrounding seabed. In particular, it is especially difficult to detect targets when sand ripples or posidonia are present. Although this conclusion was already typically accepted as truth, this study provides important quantitative evidence of its veracity.

To better understand why the detection algorithm failed to detect the targets it did, an examination of the SAS images containing the 123 missed detection opportunities was conducted. Of the 123 misses, 19 (15% of the misses) could be attributed to weak target echoes, 22 (18%) could be attributed to the target casting insufficient shadows, 31 (25%) were due to effects of



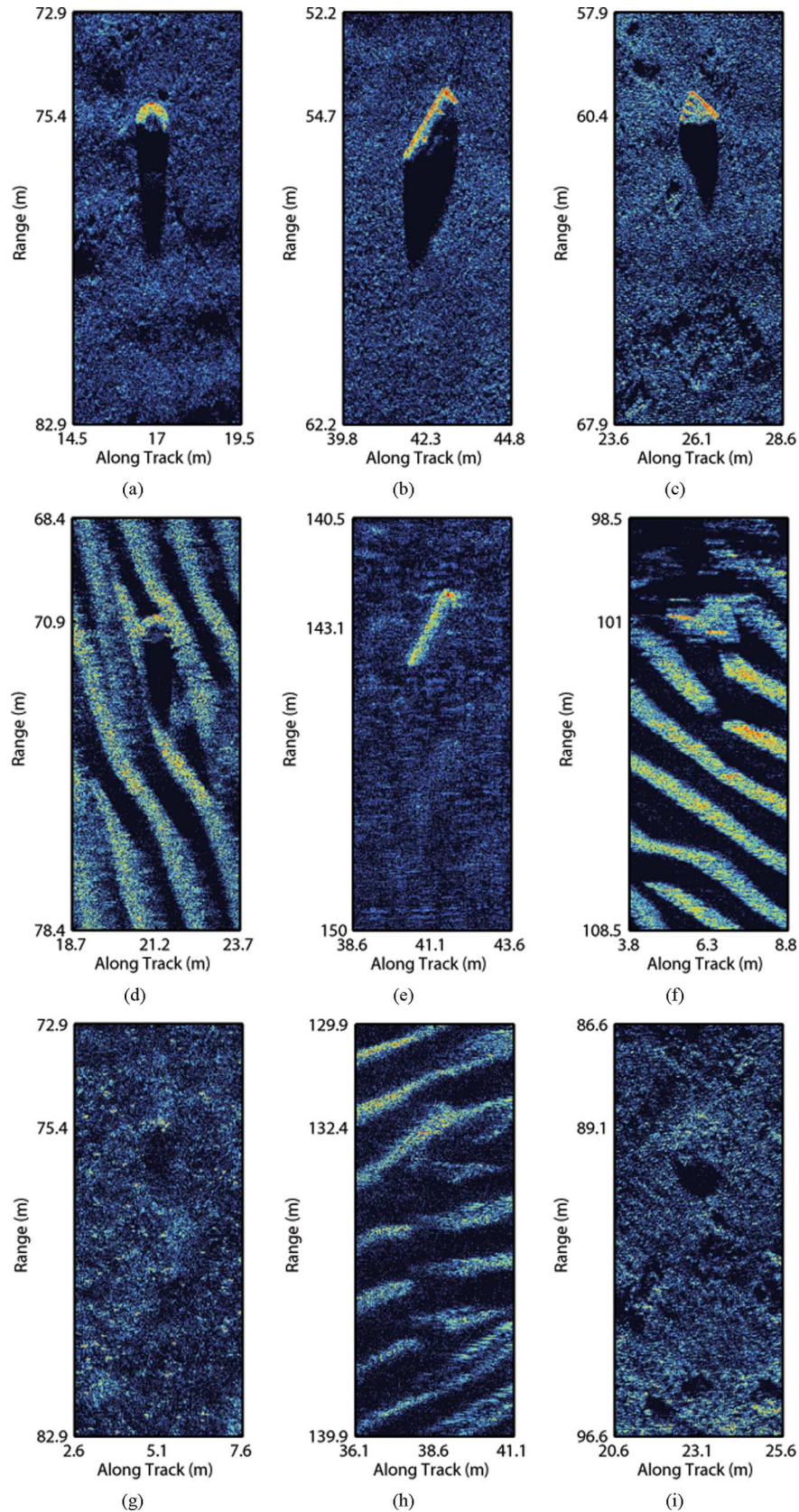


Fig. 17. Example SAS mugshots of targets considered in the study corresponding to (a)–(c) successful detections of easy cases, (d)–(f) successful detections of difficult cases, and (g)–(i) missed detections. (a) Truncated cone. (b) Cylinder. (c) Wedge. (d) Truncated cone. (e) Cylinder. (f) Wedge. (g) Truncated cone. (h) Cylinder. (i) Wedge.

sand ripples, and 51 (41%) were due to effects of posidonia. In the latter two cases, the target's echo and/or shadow would typ-

ically be significantly disrupted, obscured, or altered by the en-

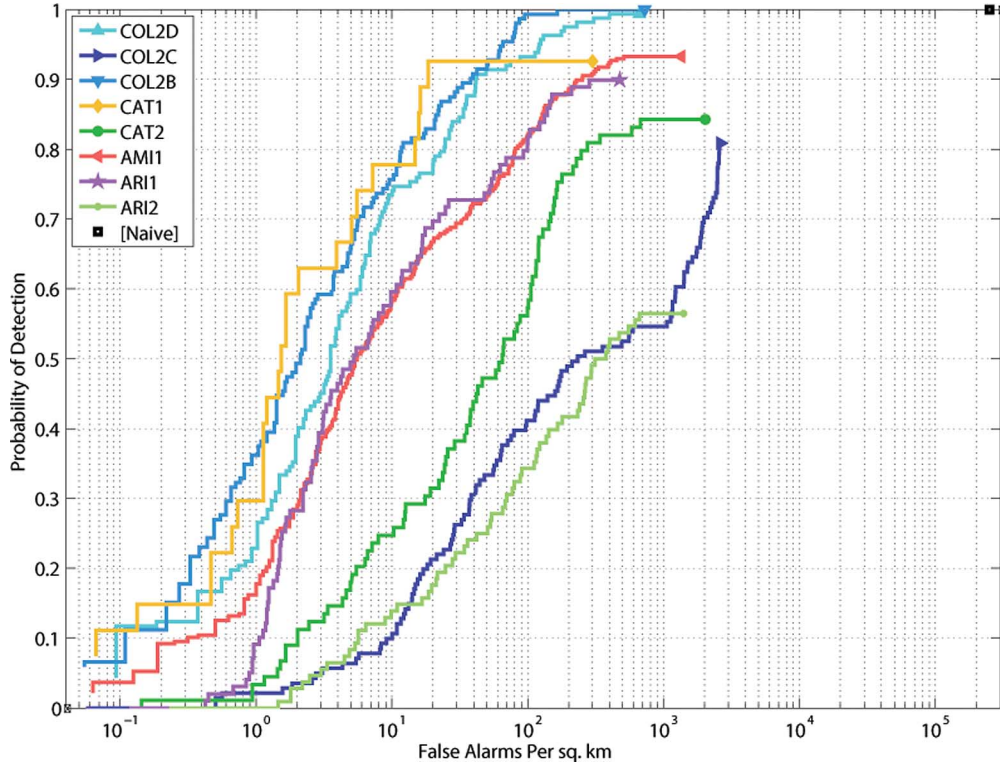


Fig. 18. Detection performance for each data set when the ordering uses the new concentration feature. Also shown is the performance of a naive detector in which a target is declared in each image at every possible position.

TABLE III  
DETECTION PERFORMANCE SUMMARY AS A FUNCTION  
OF TARGET TYPE AND SEABED

Seabed	Object	Opportunities	Detections	Probability of Detection
Benign	Cylinder	284	269	0.9472
	Truncated Cone	430	416	0.9674
	Wedge	199	187	0.9397
	[All]	<b>913</b>	<b>872</b>	<b>0.9551</b>
Ripples	Cylinder	37	29	0.7838
	Truncated Cone	42	29	0.6905
	Wedge	21	11	0.5238
	[All]	<b>100</b>	<b>69</b>	<b>0.6900</b>
Posidonia	Cylinder	0	0	—
	Truncated Cone	75	39	0.5200
	Wedge	17	2	0.1176
	[All]	<b>92</b>	<b>41</b>	<b>0.4457</b>
[All]	Cylinder	321	298	0.9283
	Truncated Cone	547	484	0.8848
	Wedge	237	200	0.8439
	[All]	<b>1105</b>	<b>982</b>	<b>0.8887</b>

vironment. Representative examples of these missed detections were shown earlier in Fig. 17(g)–(i).

To display the relative challenge of detection at different geographical sites—with varying clutter conditions—the detection performance for each data set is shown on the same plot

in Fig. 18. Rather than presenting receiver operating characteristic (ROC) curves, the results are presented as ROC-like curves, with false alarm rate replacing probability of false alarm; this enables fairer and more informative comparisons of performance at different sites. The curves in this figure are obtained by varying the detection threshold when using the new concentration feature from (15) as the final detection score. For context, also shown is the hypothetical performance of a naive detector in which a target is declared in each image at “every” possible position (viz. the “atomic unit” of an image is assumed to be a  $2\text{-m} \times 2\text{-m}$  patch of seabed for this exercise).

From Fig. 18, it can be observed that the data sets with the highest false alarm rates correspond to those sites in which boulders (with sizes comparable to targets) and other clutter objects were prevalent (cf., Table II). Detecting these types of false alarms is one tradeoff that must be made to retain the possibility of detecting new, unexpected target types. (The rejection of these sorts of natural clutter objects is a task assumed to be addressed by a later, follow-on classification stage employing a statistical learning algorithm, such as a relevance vector machine [37]. But the focus of this study is strictly detection, not sophisticated classification.) The naive detector would achieve a perfect probability of detection, but with an excessively high false alarm rate, approximately 100–1000 times higher than that of the proposed detection algorithm. This head-to-head comparison underscores the fact that the detection algorithm is indeed performing significant data reduction.

### B. Image-Quality Dependence

Next, the effects of image quality on detection performance were investigated. To this end, the detection algorithm described



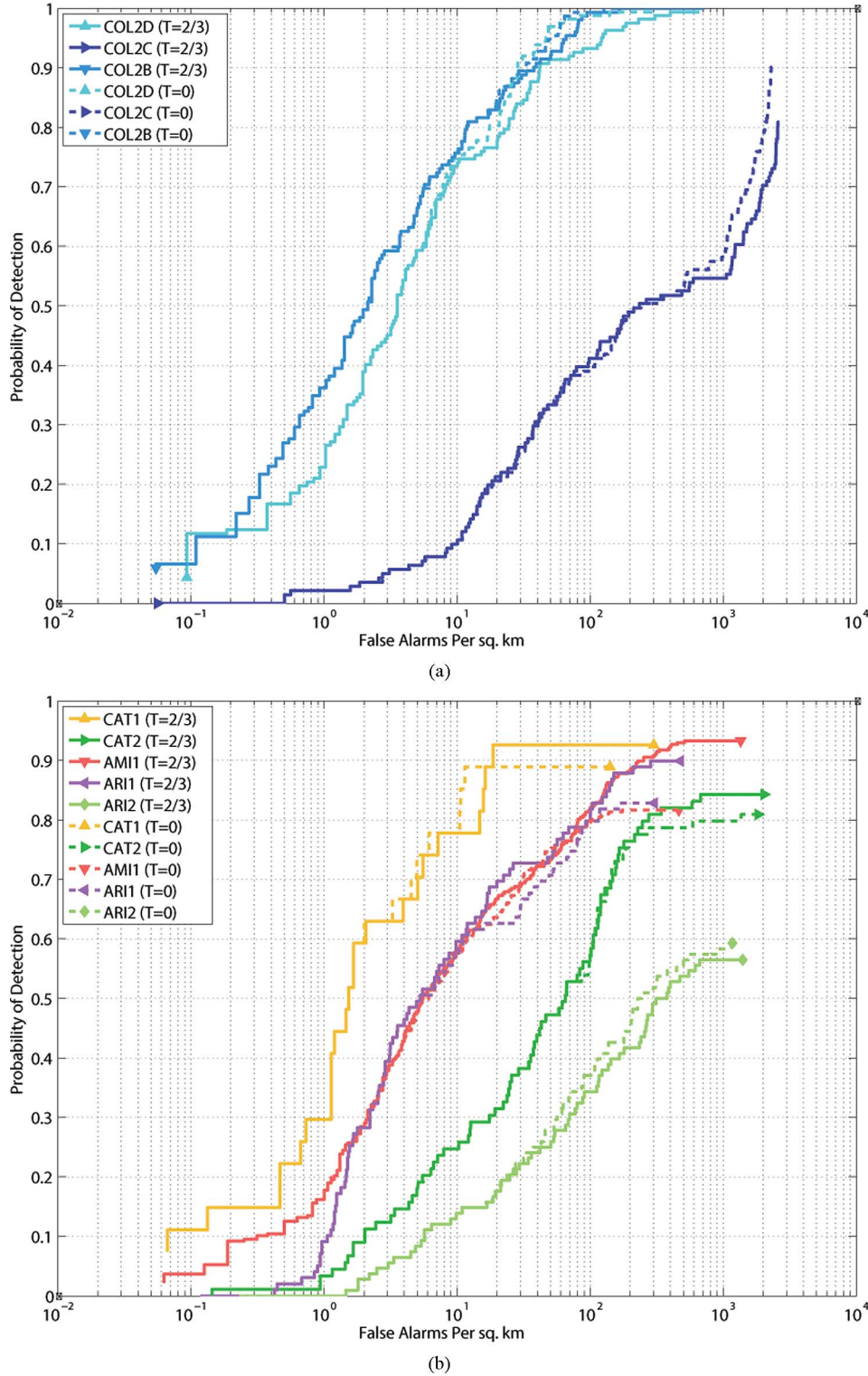


Fig. 19. Detection performance as a function of  $\tau_\rho$  for: (a) the three Colossus 2 data sets and (b) the five non-Colossus 2 data sets.

in Section II was applied to all SAS images using two different settings for  $\tau_\rho$ , the threshold below which image quality is considered poor. With the setting  $\tau_\rho = 0$ , image quality is always considered good everywhere, so it is never taken into account. In contrast, the standard setting advocated for use in the detection algorithm is  $\tau_\rho = 2/3$ , which has been observed as the point below which image quality in SAS images becomes visibly degraded. The detection performance for each data set in

terms of overall probability of detection and false alarm rate for the two  $\tau_\rho$  settings is shown in Fig. 19 (with the split across two subfigures made to improve legibility).

In general, by taking image quality into account, the probability of detection increases, as intended. (For the AMiCa data set, which particularly suffered from poor image quality at longer ranges due to the shallow-water environment, the improvement was substantial.) The tradeoff is that the false alarm

rate becomes higher as well. The amount by which the false alarm rate increases will depend on several factors, including the proportion of sonar imagery characterized by poor image quality and the clutter content in those regions. For example, in the Colossus 2–Area B data set, the vast majority of the data was of good quality, so the increase in false alarm rate from taking image quality into account was very modest.

The counterintuitive drop in detection probability in Fig. 19 from the  $\tau_p = 0$  case to the  $\tau_p = 2/3$  case for the Colossus 2–Area C and ARISE 2 data sets is due to the density-filtering step in the detection algorithm. (This was verified by running the detection algorithm with the density-filtering step deactivated; when this was done, the probabilities of detection for the  $\tau_p = 2/3$  case rose to 0.9149 and 0.6389 for the Colossus 2–Area C and ARISE 2 data sets, respectively. For the sake of consistency though, those superior detection results are not used here.) When the seabed is complex (e.g., sand ripples or posidonia) and the image quality is poor, two counteracting forces are at play in the detection algorithm: by accounting for image quality, more targets will be detected, but the weakened detection criterion will mean more false alarms are also generated by the complex seabed. The density-filtering step may then wrongly reject targets in complex seabed if they are located in high contact-density areas, effectively eliminating the gains obtained from accounting for image quality. In the Colossus 2–Area C data set, this occurred because several targets were located in sand ripples; in the ARISE 2 data set, several targets were located in posidonia.

The lesson learned from this is that the density-filtering step of the detection algorithm should be used with caution when image quality is poor in areas of complex seabed. Ideally, one would simply abstain from using poor image-quality data altogether, choosing instead to resurvey regions where image quality was poor [38] and then performing detection only on the newly collected (good quality) data. This option, however, is not always operationally feasible.

To prevent the deleterious density-filtering effect from obscuring the benefit of accounting for image quality, we next examine detection performance for targets only in *benign* seabed. Specifically, overall detection performance, across all experiments for all targets in benign seabed, as a function of *target* image quality, for cases where the detection algorithm ignores image quality (i.e., the  $\tau_p = 0$  case) or accounts for image quality (i.e., the  $\tau_p = 2/3$  case), is shown in Fig. 20.

As can be seen in the figure, detection performance is very good when the image quality is high. However, if the image quality is not taken into account (i.e., for  $\tau_p = 0$ ), performance will degrade with decreasing image quality. This is to be expected because it is known that targets will not typically have strong shadows when image quality is poor, which, in turn, makes detection less likely. Fortunately, accounting for image quality can ameliorate this situation. Fig. 20 also shows that the  $\tau_p = 2/3$  case raises the detection probability in the regime where the image quality is poor (i.e., below  $\tau_p$ ), precisely as it is intended to do. Therefore, by explicitly accounting for image quality as the detection algorithm does, the detection-performance dependence on image quality is reduced. This result is particularly important when detection must be performed but

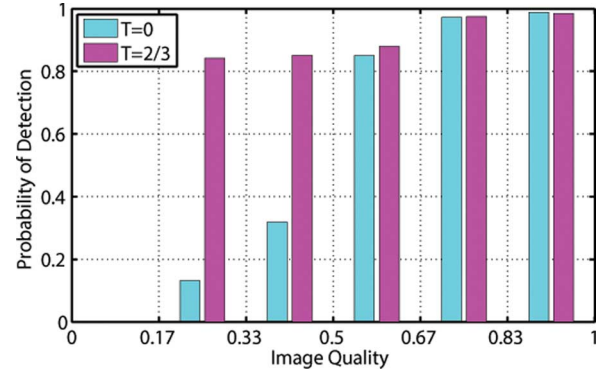


Fig. 20. Overall detection performance across all experiments for all targets in benign seabed, as a function of *target* image quality, for when image quality is ignored ( $\tau_p = 0$ ) or accounted for ( $\tau_p = 2/3$ ) by the detection algorithm. (N.B.  $\tau$  in the legend is  $\tau_p$ .)

only poor quality data are available, with no opportunity to resurvey at sea to collect better quality data.

### C. Value of New Detection Score

Next, we wish to establish the utility of employing the new concentration feature from (15) as the final detection score for contacts. To this end, a comparison of target detection performance (with the standard  $\tau_p = 2/3$  setting) is made when the final detection score is the new concentration feature and when it is the old feature based on echo strength as in [18]. (In this context, the final detection score can be viewed as the “prediction” of a trivial single-feature classifier.) The detection performance for the eight data sets is shown in Fig. 21 (with the split across two subfigures made to improve legibility).

It can be seen in the figure that the case using the new concentration feature as the final detection score discriminates targets from clutter better than the case employing the old echo strength feature. In particular, this improved ordering of contacts can be observed clearly at the lowest false alarm rates, where the probability of detection rises much faster with the new score. In fact, accurate ordering of contacts in this regime, which corresponds to the highest (i.e., most mine-like) scores, is very important for *in situ* contact reinspection.

Resource constraints at sea, such as AUV battery life, mean that typically only a subset of detected objects can be reinspected. Therefore, it is extremely beneficial that the contacts assigned the highest score are indeed targets, as this means the few additional views that *are* acquired will be of high-value objects. To demonstrate the value of the new feature in this regard, Table IV shows the number of targets that are among the top 100 contacts with the highest final detection scores, when the score is based on the new concentration feature or the old echo-strength feature. As can be observed from the table, the new feature is demonstrably better at assigning the highest detection scores to targets.

To more quantitatively demonstrate the overall superiority of the new concentration score for ordering the detected contacts, the area under the ROC curve (AUC) [39] is calculated when using the new concentration score and also when using the old echo-strength score. The AUC provides a scalar summary measure of a ROC curve; an AUC of unity would indicate that all targets were detected without incurring a single false alarm (i.e.,

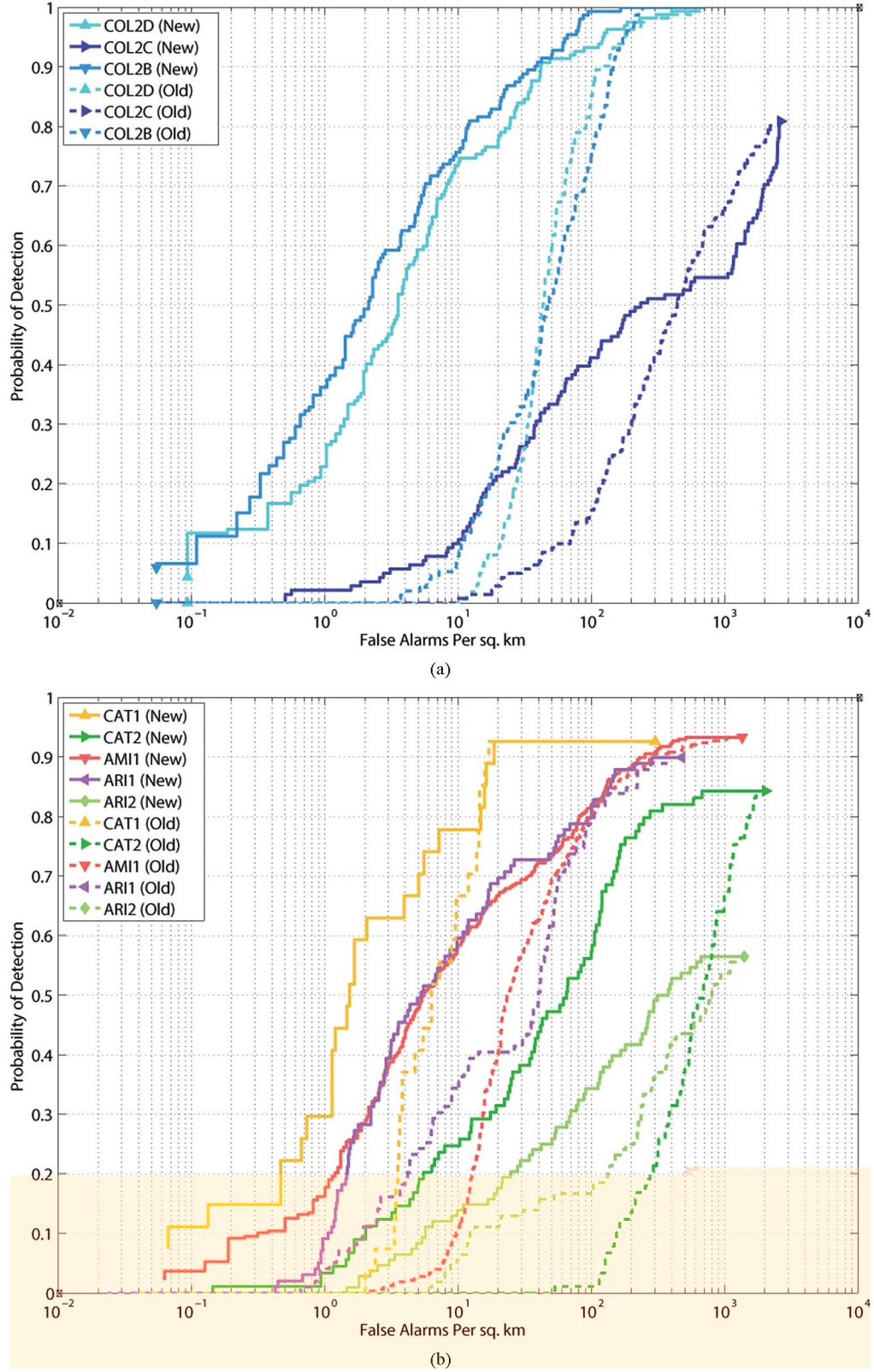


Fig. 21. Detection performance as a function of final detection score used for: (a) the three Colossus 2 data sets and (b) the five non-Colossus 2 data sets. “New” denotes the concentration feature; “old” denotes the echo strength feature.

each target had a higher detection score than every clutter object). (For the AUC calculation, the ROC-like curves in Fig. 21 must first be converted to proper ROC curves. To ensure that the maximum possible AUC value is unity for each data set, the probabilities of detection and false alarm are scaled to range

from zero to unity for each data set.) The AUC for the two detection scores considered are shown for each data set in Table V.

As can be seen from the table, the new feature is typically a much better predictor for discriminating targets from clutter, as evidenced by higher AUC values with the new feature. The



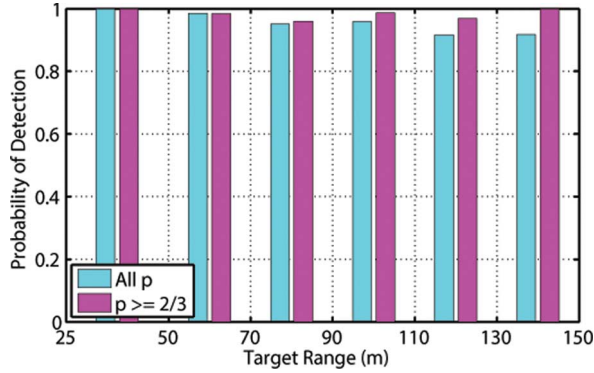


Fig. 22. Overall detection performance across all data sets for all target types in benign seabed, as a function of target range, when considering only good image-quality target opportunities ( $\rho \geq 2/3$ ) or all opportunities (all  $\rho$ ).

TABLE IV  
NUMBER OF TARGETS AMONG THE SET OF 100 CONTACTS WITH THE HIGHEST FINAL DETECTION SCORES, WHEN BASED ON DIFFERENT FEATURES

Data Set	Feature as Final Detection Score	
	Echo Strength (Old)	Concentration (New)
COL2D	0	<b>87</b>
COL2B	6	<b>88</b>
COL2C	0	<b>10</b>
CAT1	19	<b>27</b>
CAT2	1	<b>26</b>
AMI1	48	<b>97</b>
ARI1	18	<b>37</b>
ARI2	0	<b>56</b>

one exception to this trend occurs with the Colossus 2–Area C data set, where several targets were placed in sand ripples. This result can be explained by the fact that the concentration feature calculation relies on a contact’s full mugshot, so a background with sand ripples—and, in particular, the strong shadows and strong echoes induced by them—will necessarily impact the feature value. However, it should be noted that the feature was still useful in posidonia, as evidenced by the improved AUC for the ARISE 2 data set. Thus, the new feature is very valuable provided sand ripples are not present.

#### D. Target Aspect and Range Dependence

Having already established the impact of the seabed environment on detection, the following assessment considers only cases in which the target was in neither sand ripples nor posidonia. This allows us to remove one variable from consideration and to better isolate the quantities that impact detection performance in benign seabeds.

Detection performance as a function of target range from the sensor is shown in Fig. 22 for all targets in benign seabed across all data sets (using the standard  $\tau_\rho = 2/3$  setting). One case shows detection performance for all target detection opportunities; another case shows performance if only good image-quality opportunities are considered (i.e., excluding targets in poor image-quality regions, where  $\rho < 2/3$ ). Showing these two cases together highlights the fact that the slight decrease in

TABLE V  
AUC WHEN ORDERING DETECTIONS WITH DIFFERENT SCORES

Data Set	Feature as Final Detection Score	
	Echo Strength (Old)	Concentration (New)
COL2D	0.9051	<b>0.9692</b>
COL2B	0.9091	<b>0.9841</b>
COL2C	<b>0.8029</b>	0.7451
CAT1	0.9742	<b>0.9863</b>
CAT2	0.6825	<b>0.9599</b>
AMI1	0.9542	<b>0.9718</b>
ARI1	0.8996	<b>0.9451</b>
ARI2	0.7752	<b>0.9057</b>

TABLE VI  
DETECTION PERFORMANCE AS A FUNCTION OF ASPECT FOR CYLINDERS IN BENIGN SEABED

Target Section	Range of Aspects	Opportunities	Detections	Probability of Detection
Front	[315°, 45°)	74	71	0.9595
Right Side	[45°, 135°)	67	64	0.9552
Rear	[135°, 225°)	75	72	0.9600
Left Side	[225°, 315°)	68	62	0.9118

TABLE VII  
DETECTION PERFORMANCE AS A FUNCTION OF ASPECT FOR WEDGES IN BENIGN SEABED

Target Section	Range of Aspects	Opportunities	Detections	Probability of Detection
Front	[315°, 45°)	53	51	0.9623
Right Side	[45°, 135°)	46	42	0.9130
Rear	[135°, 225°)	55	53	0.9636
Left Side	[225°, 315°)	45	41	0.9111

detection performance at long range in the former case is in fact due to poor image quality, not range itself. If image quality is sufficient, performance exhibits no dependence on range, as can be readily seen with the results of the  $\rho \geq 2/3$  case. Indeed, this result is to be expected because SAS imagery has constant resolution in range [2].

Next, detection performance as a function of target orientation is shown for cylinders and wedges in benign seabed, across all data sets, in Tables VI and VII, respectively.<sup>4</sup> (The aspects are divided into four bins, each spanning 90° and corresponding to the front, sides, and rear of the object; 0° corresponds to the front of the object.) As can be seen from the tables, detection performance for the targets does not exhibit strong aspect dependence. The result that cylinders viewed at endfire were not

<sup>4</sup>The (essentially) rotationally symmetric nature of the truncated cones precludes the accurate association of a given aspect (i.e., orientation) of a truncated cone from one sea experiment with that from another sea experiment. This means that it is not possible to create a common reference frame for purposes of examining aspect-dependent detection performance. Therefore, overall detection performance (from combining data from multiple sea experiments) as a function of aspect is not shown for the truncated cone.

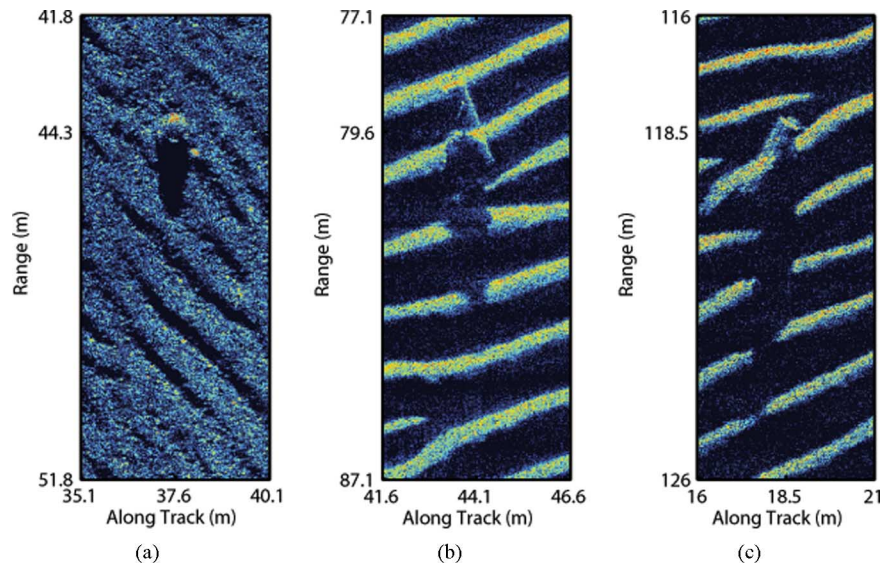


Fig. 23. Example SAS mugshots of (a) a truncated cone in sand ripples oriented at  $130^\circ/310^\circ$ , (b) a cylinder oriented nearly orthogonal to sand ripples oriented at  $115^\circ/295^\circ$ , and (c) a cylinder oriented nearly parallel to sand ripples oriented at  $175^\circ/355^\circ$ . (The reference frames are such that endfire for each cylinder is  $0^\circ$ .)

TABLE VIII  
DETECTION PERFORMANCE AS A FUNCTION OF TARGET ASPECT FOR A TRUNCATED CONE IN SAND RIPPLES ORIENTED AT  $130^\circ/310^\circ$

Range of Aspects	Sonar Look Direction	Opportunities	Detections	Probability of Detection	95% Confidence Interval
$[85^\circ, 175^\circ)$	Along Ripples	7	6	0.8571	[0.4869, 0.9743]
$[175^\circ, 265^\circ)$	Across Ripples	3	1	0.3333	[0.0615, 0.7923]
$[265^\circ, 355^\circ)$	Along Ripples	8	6	0.7500	[0.4093, 0.9285]
$[355^\circ, 85^\circ)$	Across Ripples	2	1	0.5000	[0.0945, 0.9055]

more challenging to detect than cylinders viewed at broadside can be attributed to the fact that the number of pixels on target in both cases was sufficiently large. Only when there are very few pixels on target (e.g., with a lower resolution sonar) should detection performance be expected to vary with cylinder orientation.

#### E. Impact of Sand-Ripple Orientation

It was observed that the ability to detect targets has a strong dependence on the presence of sand ripples. But, in particular, the *orientation* at which the sand ripples are surveyed has a dramatic impact on performance as well. (The orientation of a sand ripple is defined here to be the direction along a ripple crest.) This phenomenon is illustrated here with three different targets that were located in sand ripple fields; example SAS mugshots (resulting from successful detections) of the objects are shown in Fig. 23.

The dependence on sand-ripple orientation can first be observed in the case of a rotationally symmetric truncated cone in sand ripples, where detection performance is still aspect dependent. Detection performance for this target, corresponding to the SAS mugshot in Fig. 23(a), is shown in Table VIII. Because of the small sample sizes, the Agresti–Coull 95% confidence intervals [40] are also shown.

In this case, detection performance is much better when the sonar viewing aspect is looking along a ripple crest (or trough). Therefore, when sand ripples are present, it is recommended that

one surveys such that the sonar looks along ripple crests (and troughs) rather than across them. In this way, one minimizes the chance that targets will be obscured by the ripples, either from target and ripple highlights melding or from ripple shadow concealment.

Another interesting phenomenon regarding detection performance in ripples can be observed from nonrotationally symmetric targets. In the Colossus 2 sea experiment at Area C, two cylinders were located in the same sand-ripple field, but at different aspects with respect to the ripples. One target [corresponding to the SAS mugshot in Fig. 23(b)] was oriented nearly orthogonal to the ripple crests, whereas the other target [corresponding to the SAS mugshot in Fig. 23(c)] was oriented nearly parallel to the ripple crests. Detection performance as a function of target aspect for these two targets is shown in Tables IX and X, respectively. (Endfire for each cylinder is  $0^\circ$ .) As before, the Agresti–Coull 95% confidence intervals are also shown.

As can be observed in Table IX, the arrangement in which the target was oriented nearly orthogonal to the ripple crests produced no significant aspect dependence for detection. In contrast, in Table X, strong aspect dependence for detection was exhibited for the target that was oriented nearly parallel to the ripple crests. (In fact, given the physical cylinder-ripple arrangement, the detection probabilities for the cylinder were higher for the orientations closer to *endfire*.) While the admittedly small sample sizes (of detection opportunities) preclude drawing firm conclusions, these anecdotal results do suggest that the relative



TABLE IX

DETECTION PERFORMANCE AS A FUNCTION OF TARGET ASPECT FOR A CYLINDER ORIENTED NEARLY ORTHOGONAL TO SAND RIPPLES ORIENTED AT  $115^\circ/295^\circ$ 

Target Section	Range of Aspects	Opportunities	Detections	Probability of Detection	95% Confidence Interval
Front	$[315^\circ, 45^\circ]$	4	3	0.75	[0.3006, 0.9544]
Right Side	$[45^\circ, 135^\circ]$	6	4	0.67	[0.3000, 0.9032]
Rear	$[135^\circ, 225^\circ]$	4	3	0.75	[0.3006, 0.9544]
Left Side	$[225^\circ, 315^\circ]$	5	4	0.80	[0.3755, 0.9638]

TABLE X

DETECTION PERFORMANCE AS A FUNCTION OF TARGET ASPECT FOR A CYLINDER ORIENTED NEARLY PARALLEL TO SAND RIPPLES ORIENTED AT  $175^\circ/355^\circ$ 

Target Section	Range of Aspects	Opportunities	Detections	Probability of Detection	95% Confidence Interval
Front	$[315^\circ, 45^\circ]$	5	5	1.0	[0.5655, 1.0]
Right Side	$[45^\circ, 135^\circ]$	4	3	0.75	[0.3006, 0.9544]
Rear	$[135^\circ, 225^\circ]$	5	5	1.0	[0.5655, 1.0]
Left Side	$[225^\circ, 315^\circ]$	4	2	0.50	[0.1500, 0.8500]

orientation of targets to sand ripples indeed has an impact on detection ability. The most challenging case appears to be when the principal axis of the target and a ripple crest align. This result makes intuitive sense because it is in such arrangements that the highlights of the target and ripples would coincide, effectively masking the target response.

#### F. Generalizing to New Targets

In addition to the three aforementioned target shapes, other objects were also deployed at sea during the experiments; among these were two mine-size rocks and two unconventional man-made objects—a household washing machine and a large gym bag—each meant to simulate a potential improvised explosive device (IED). Photographs and example SAS images of these objects are shown in Fig. 24.

The detection performance for these nontarget objects—rocks and pseudo-IEDs—is included here to demonstrate the generalization ability of the algorithm to detect new or unexpected target shapes of sizes comparable to conventional targets. This functionality is important because it is always possible to encounter new or unexpected targets at sea. A detection algorithm that instead rigidly trained on specific target types would likely miss targets that were unknown *a priori* or absent from the training set. (Nevertheless, for the target detection performance study presented earlier, these nontarget objects were treated as false alarms, and they were not counted in the target detection opportunities in Table II.) The detection performance for these nontarget objects in benign seabed is shown in Table XI, where it can be seen that the results are comparable to those for the target objects (cf., Table III).

## V. DISCUSSION

The overarching motivation for the proposed detection algorithm was to overcome the limitations of the most popular existing methods and enable near-real-time detection onboard an AUV with limited processing power. A balance of rigor

and flexibility was embedded in the detection algorithm by exploiting definitions (shadows, ripple fields), by accounting for geometry (range-dependent shadow lengths and grazing angle effects), by incorporating extensive domain-specific knowledge (object sizes for template sizes and thresholds), and by adapting to *in situ* environmental and data conditions (background reverberation level, sand ripples, image quality). Vitaly, the overall algorithm was made fast by exploiting integral-image representations and by employing a cascaded architecture that operated on progressively smaller portions of the image at each step. The detection algorithm can also be used for lower resolution *sidescan* sonar images, provided the image-quality check in Section II-B4 is elided.

Several interesting results have been revealed by the detection performance study that was conducted. It was observed that image quality is indeed an important factor for detecting targets in SAS images. Moreover, it was also demonstrated that the particular detection algorithm employed was able to successfully take image quality into account and, in the process, effectively reduce its impact on target-detection performance.

The study also supplied quantitative values for detection probabilities and false alarm rates over a huge set of real SAS data collected at sea without resorting to injected targets [41] or other synthetic, less realistic data. It was demonstrated that a very high probability of detection could be achieved in benign seabeds with reasonable false alarm rates. This can largely be attributed to the fact that the high-resolution SAS imagery permits many pixels on target. The difficulty of performing detection in posidonia and sand ripples was also highlighted. Additionally, based on detection results for nontarget objects of sizes comparable to conventional targets, the detection algorithm can detect new, unexpected target shapes, as it does not train on a specific set of targets (nor does it require *any* training data, in fact).

Detection performance was found not to depend strongly on the range of a target from the sensor, which is expected

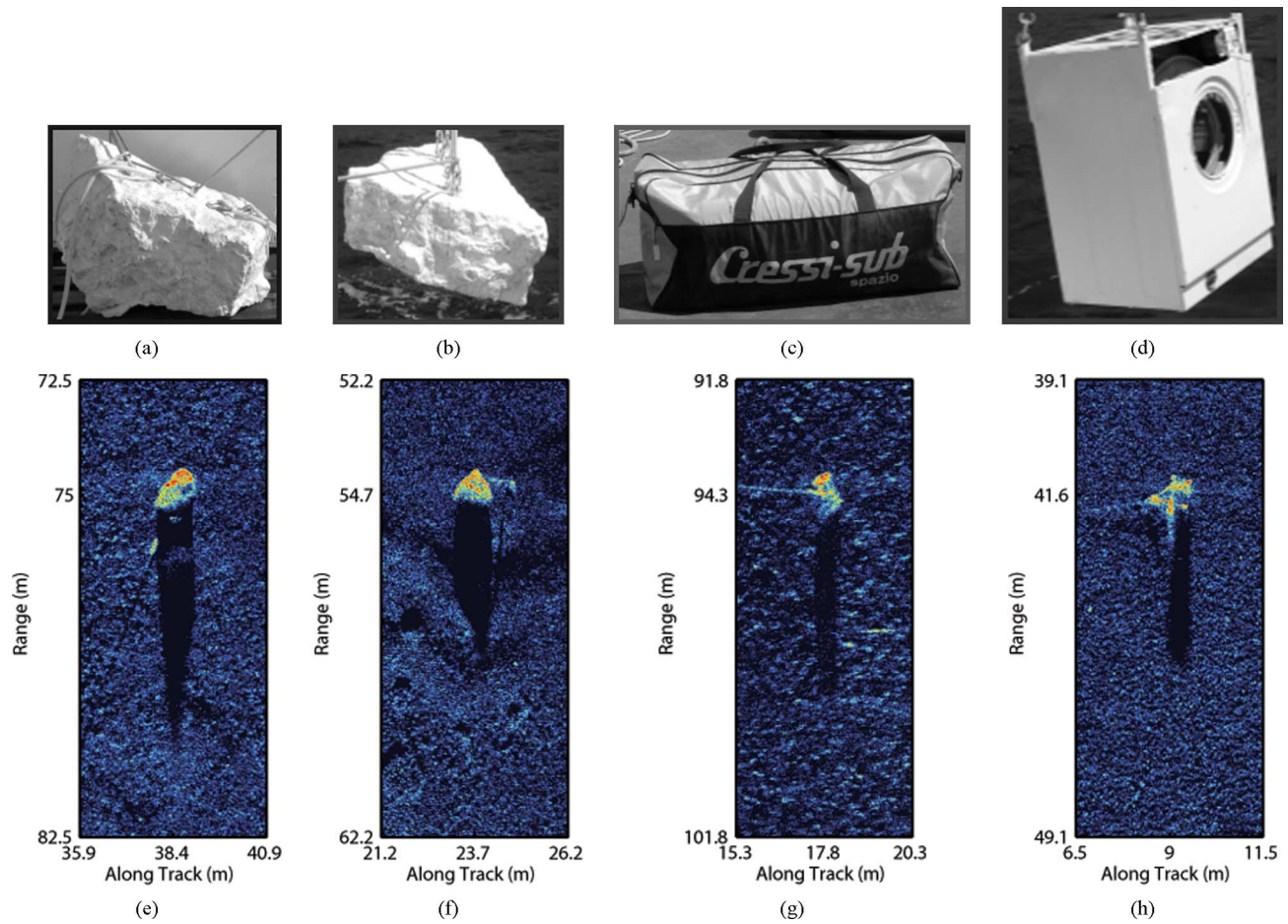


Fig. 24. Nontarget objects considered in the study, shown in (a)–(d) photographs and (e)–(h) example SAS mugshots, corresponding to successful detections. The captions indicate the object codes. (a) VR01. (b) FR01. (c) IED01. (d) HM01. (e) VR01. (f) FR01. (g) IED01. (h) HM01.

since SAS imagery has constant resolution in range [2], unlike sidescan sonar imagery. It was also observed that neither target type nor target aspect played a major role in detection performance. The one exception to this finding was for targets in sand-ripple fields, where the relative orientation between the target and the ripples did influence detection performance.

The results of the study revealed that the environment on the seafloor has a significant impact on detection ability. In particular, the presence of posidonia and sand ripples can make detection particularly challenging. To combat this phenomenon, it may be useful to create alternative detection algorithms tailored specially to these conditions. Although the detection algorithm employed already automatically adapts to the local properties of the site, additional special allowances may be required for the case of posidonia.

Complications arise for detection when data are characterized by poor image quality and when complex seabeds are involved. These issues can be mitigated by intelligently resurveying in the following ways. To guard against poor image quality, the tracks of an AUV survey can be adapted to ensure that good quality data are collected over the entire area of interest [38]. Regarding sand ripples, once they are detected (as a subcomponent of the detection algorithm), the principal orientation of the AUV survey can be adapted so that the tracks are oriented parallel to the direction in which the ripple “waves” propagate.

TABLE XI  
DETECTION PERFORMANCE SUMMARY OF NONTARGET  
OBJECTS IN BENIGN SEABED

Object Code	Object Description	Opportunities	Detections	Probability of Detection
VR01	Rock	110	107	0.9727
FR01	Rock	67	60	0.8955
IED01	Gym Bag	20	17	0.8500
HM01	Washing Machine	25	24	0.9600

To prevent viewing targets at only unfavorable aspects (*vis-à-vis* classification), an additional survey that exhibits substantial aspect diversity (such as a  $90^\circ$  orientation difference) can be executed [42]. For this contact reinspection, the new detection score based on the highlight and shadow concentration features was shown to be a powerful “single-feature classifier” for ordering contacts as mine-like. (Additionally, the concentration feature should be a salient feature for a subsequent “proper” classification stage.) As a result, the limited resources possessed at sea can be efficiently allocated to enable the acquisition of additional views for the subset of most suspicious objects.

## VI. CONCLUSION

A new unsupervised target detection algorithm for SAS imagery was proposed and evaluated. The method achieves good performance partially by capitalizing on the fact that the high-quality, high-resolution SAS imagery permits many pixels on target. The performance of the algorithm was demonstrated via a large-scale study using real SAS imagery collected at sea by the MUSCLE AUV in different conditions. Numerous quantitative results regarding detection performance as a function of target type, aspect, range, image quality, seabed environment, and geographical site were obtained. The impacts of image quality, seabed environment, and sand-ripple orientation on detection capabilities were also highlighted.

Because the SAS data sets studied here have been widely distributed to many researchers, these detection performance results can act as a benchmark of sorts against which other detection algorithms can be compared in the future. The results can also be utilized for the design of performance planning and evaluation tools for target detection with SAS-equipped AUVs.

More immediately, the results can be used to support and guide algorithm development for optimizing adaptive AUV surveys using through-the-sensor data, for which research has already begun and is ongoing at CMRE [38], [42]. In fact, during a recent sea experiment in 2013 with the MUSCLE AUV, CMRE successfully demonstrated *in situ* target detection (using the proposed algorithm) and immediate target reinspection. Future work will continue this strand of research directed toward enabling fully autonomous MCM operations.

## REFERENCES

- [1] B. Walters, J. French, and M. Barnes, "Modeling the effects of crew size and crew fatigue on the control of tactical unmanned aerial vehicles (TUAVs)," in *Proc. Conf. Simul.*, 2000, pp. 920–924.
- [2] M. Hayes and P. Gough, "Broad-band synthetic aperture sonar," *IEEE J. Ocean. Eng.*, vol. 17, no. 1, pp. 80–94, Jan. 1992.
- [3] J. Johnson, "Analysis of image forming systems," in *Proc. Image Intensifier Symp.*, 1958, pp. 244–273.
- [4] G. Dobeck, J. Hyland, and L. Smedley, "Automated detection/classification of seamounts in sonar imagery," *Proc. SPIE—Int. Soc. Opt. Eng.*, vol. 3079, pp. 90–110, 1997.
- [5] S. Reed, Y. Petillot, and J. Bell, "An automatic approach to the detection and extraction of mine features in sidescan sonar," *IEEE J. Ocean. Eng.*, vol. 28, no. 1, pp. 90–105, Jan. 2003.
- [6] J. Fawcett, A. Crawford, D. Hopkin, V. Myers, and B. Zerr, "Computer-aided detection of targets from the CITADEL Trial Klein sonar data," Defence R&D Canada—Atlantic, Dartmouth, NS, Canada, Tech. Rep. TM 2006-115, 2006.
- [7] F. Maussang, J. Chanussot, A. Hétet, and M. Amate, "Mean-standard deviation representation of sonar images for echo detection: Application to SAS images," *IEEE J. Ocean. Eng.*, vol. 32, no. 4, pp. 956–970, Oct. 2007.
- [8] J. Groen, E. Coiras, and D. Williams, "Detection rate statistics in synthetic aperture sonar images," in *Proc. Int. Conf. Underwater Acoust. Meas.*, 2009, pp. 367–374.
- [9] J. Sawas, Y. Petillot, and Y. Pailhas, "Cascade of boosted classifiers for rapid detection of underwater objects," in *Proc. Eur. Conf. Underwater Acoust.*, 2010.
- [10] J. Tucker and M. Azimi-Sadjadi, "Coherence-based underwater target detection from multiple disparate sonar platforms," *IEEE J. Ocean. Eng.*, vol. 36, no. 1, pp. 37–51, Jan. 2011.
- [11] S. Williams, O. Pizarro, M. How, D. Mercer, G. Powell, J. Marshall, and R. Hanlon, "Surveying nocturnal cuttlefish camouflage behaviour using an AUV," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2009, pp. 214–219.
- [12] Z. Reut, N. Pace, and M. Heaton, "Computer classification of seabeds by sonar," *Nature*, vol. 314, pp. 426–428, 1985.
- [13] H. Singh, J. Adams, D. Mindell, and B. Foley, "Imaging underwater for archaeology," *J. Field Archaeol.*, vol. 27, no. 3, pp. 319–328, 2000.
- [14] Y. Petillot, S. Reed, and J. Bell, "Real time AUV pipeline detection and tracking using side scan sonar and multi-beam echo-sounder," in *Proc. IEEE OCEANS Conf.*, 2002, pp. 217–222.
- [15] Y. Zhang, X. Liao, and L. Carin, "Detection of buried targets via active selection of labeled data: Application to sensing subsurface UXO," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 11, pp. 2535–2543, Nov. 2004.
- [16] A. Dabbagh, K. Al-Hinai, and M. Khan, "Detection of sand-covered geologic features in the Arabian Peninsula using SIR-C/X-SAR data," *Remote Sens. Environ.*, vol. 59, no. 2, pp. 375–382, 1997.
- [17] D. Williams and J. Groen, "A fast physics-based, environmentally adaptive underwater object detection algorithm," in *Proc. IEEE OCEANS Conf.*, 2011, DOI: 10.1109/Oceans-Spain.2011.6003424.
- [18] D. Williams, "On adaptive underwater object detection," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2011, pp. 4741–4748.
- [19] P. Viola and M. Jones, "Robust real-time object detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.
- [20] M. Doherty, J. Landowski, P. Maynard, G. Uber, D. Fries, and F. Maltz, "Side scan sonar object classification algorithms," in *Proc. 6th Int. Symp. Unmanned Untethered Submersible Technol.*, 1989, pp. 417–424.
- [21] G. Dobeck, "Fusing sonar images for mine detection and classification," in *Proc. AeroSense*, 1999, pp. 602–614.
- [22] G. Widmer and M. Kubat, "Learning in the presence of concept drift and hidden contexts," *Mach. Learn.*, vol. 23, pp. 69–101, 1996.
- [23] B. Zadrozny, "Learning and evaluating classifiers under sample selection bias," in *Proc. 21st Int. Conf. Mach. Learn.*, 2004, DOI: 10.1145/1015330.1015425.
- [24] H. Cox and D. Pace, "A fast normalizer," in *Proc. 13th Asilomar Conf. Signals Syst. Comput.*, 1996, vol. 1, pp. 459–463.
- [25] W. Struzinski and E. Lowe, "A performance comparison of four noise background normalization schemes proposed for signal detection systems," *J. Acoust. Soc. Amer.*, vol. 76, no. 6, pp. 1738–1742, 1984.
- [26] R. Hansen, H. Callow, T. Sæbø, and S. Synnes, "Challenges in seafloor imaging and mapping with synthetic aperture sonar," in *Proc. Eur. Conf. Synthetic Aperture Radar*, 2010.
- [27] L. Wang, G. Davies, A. Bellettini, and M. Pinto, "Multipath effect on DPCA micronavigation of a synthetic aperture sonar," in *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance*, N. Pace and F. Jensen, Eds. Norwell, MA, USA: Kluwer, 2002, pp. 465–472.
- [28] A. Bellettini and M. Pinto, "Theoretical accuracy of synthetic aperture sonar micronavigation using a displaced phase-center antenna," *IEEE J. Ocean. Eng.*, vol. 27, no. 4, pp. 780–789, Oct. 2002.
- [29] S. Synnes, H. Callow, R. Hansen, and T. Sæbø, "Multipass coherent processing on synthetic aperture sonar data," in *Proc. Eur. Conf. Underwater Acoust.*, 2010.
- [30] S. Synnes, R. Hansen, and T. Sæbø, "Assessment of shallow water performance using interferometric sonar coherence," in *Proc. Int. Conf. Underwater Acoust. Meas.*, 2009.
- [31] P. Wiberg and C. Harris, "Ripple geometry in wave-dominated environments," *J. Geophys. Res.*, vol. 99, pp. 775–789, 1994.
- [32] D. Hanes, V. Alymov, and Y. Chang, "Wave-formed sand ripples at Duck, North Carolina," *J. Geophys. Res.*, vol. 106, pp. 22 575–22 592, 2001.
- [33] E. Reffet, S. C. du Pont, P. Hersen, and S. Douady, "Formation and stability of transverse and longitudinal sand dunes," *Geology*, vol. 38, no. 6, pp. 491–494, 2010.
- [34] A. Lyons, D. Abraham, and S. Johnson, "Modeling the effect of seafloor ripples on synthetic aperture sonar speckle statistics," *IEEE J. Ocean. Eng.*, vol. 35, no. 2, pp. 242–249, Apr. 2010.
- [35] L. Hellequin, J. Boucher, and X. Lurton, "Processing of high-frequency multibeam echo sounder data for seafloor characterization," *IEEE J. Ocean. Eng.*, vol. 28, no. 1, pp. 78–89, Jan. 2003.
- [36] R. Fandos, L. Sadamori, and A. Zoubir, "High quality segmentation of synthetic aperture sonar images using the min-cut/max-flow algorithm," in *Proc. Eur. Signal Process. Conf.*, 2011, pp. 51–55.
- [37] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.
- [38] D. Williams, A. Vermeij, F. Baralli, J. Groen, and W. Fox, "In situ AUV survey adaptation using through-the-sensor sonar data," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2012, pp. 2525–2528.
- [39] J. Hanley and B. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, pp. 29–36, 1982.

- [40] A. Agresti and B. Coull, "Approximate is better than "exact" for interval estimation of binomial proportions," *Amer. Stat.*, vol. 52, no. 2, pp. 119–126, 1998.
- [41] E. Coiras, P. Mignotte, Y. Petillot, J. Bell, and K. Lebart, "Supervised target detection and classification by training on augmented reality data," *IET Radar Sonar Navig.*, vol. 1, no. 1, pp. 83–90, Jan. 2007.
- [42] M. Couillard, J. Fawcett, and M. Davison, "Optimizing constrained search patterns for remote mine-hunting vehicles," *IEEE J. Ocean. Eng.*, vol. 37, no. 1, pp. 75–84, Jan. 2012.



**David P. Williams** received the B.S.E. (*magna cum laude*), M.S., and Ph.D. degrees in electrical and computer engineering from Duke University, Durham, NC, USA, in 2002, 2003, and 2006, respectively.

Since 2007, he has been with the Centre for Maritime Research & Experimentation, NATO Science and Technology Organization (formerly NATO Undersea Research Centre), La Spezia, Italy. His research interests lie in the fields of machine learning, pattern recognition, and mine countermeasures.

Dr. Williams was the recipient of a James B. Duke Graduate Fellowship and a National Defense Science and Engineering Graduate Fellowship.