A Screening Application for Image Data Collected by an Acoustic Lens Sonar

by J.B. Hsieh, J.I. Olsonbaker, and W.L.J. Fox

Technical Report
APL-UW TR 0502
November 2005

Applied Physics Laboratory University of Washington
1013 NE 40th Street Seattle, Washington 98105-6698

ONR N00014-01-G-0460, Delivery Order 0005

Acknowledgments

We wish to thank the Office of Naval Research (program officer Dr. Charles Loeffler) for support of this research. We also appreciate the use of the data provided by Mr. Roger Stokey of the Woods Hole Oceanographic Institution, Woods Hole, MA.

Abstract

High-resolution, forward-looking sonar systems are critical tools on autonomous underwater vehicle (AUV) platforms where they are used to detect, classify, and localize objects. Data collected by these sensors can be processed automatically and used for navigation (object avoidance) or detailed object examination and discrimination tasks. Once an object is detected, size and shape parameters can be estimated based on the acoustic shadow cast behind the object when ensonified by such sensors. This report discusses the development of adaptive algorithms for highlight localization and shadow segmentation from 1.8-MHz data collected by an acoustic lens imaging sonar—DIDSON (Dual-frequency IDentification SONar). The sonar was mounted in a forward-looking configuration on a REMUS AUV. Algorithms screen large data sets for interesting frames; they demonstrate considerable effectiveness for target detection when the target images are central to the data frame and where there is little platform motion.

Table of Contents

1. Introduction 1
Background 1
Current Work 2
2. DIDSON Description 3
3. Development and Analysis 6
Normalization 6
Motion Correction 9
Background Estimation 11
Highlight Localization 12
Shadow Segmentation 13
Matched Highlights and Shadows 16
4. Data Sets and Results 18
Data 18
Results of Data Analysis 19
Target Size Estimation 24
5. Conclusions 26
6. References 27
7. Appendix 28

1. Introduction

The long-term objective of this work is to develop algorithms that will enable AUVs to make autonomous identification of mines in the field. This report describes detections of potential targets using methods of highlight localization and shadow segmentation.

Imaging sonars collect large quantities of data [up to 21 frames per second (fps), but usually at 3 fps] that are downloaded and viewed by operators upon mission completion. The algorithms developed and described here can be used to screen large data sets for snippets of interesting frames for an operator to view. Filtering out a majority of the uninteresting data reduces the workload of an operator who may have hours of data to analyze. Our algorithms applied to the data bring all potential targets to the operator's attention. We used preliminary classification, via estimates of width and height, to reduce the quantity of spurious detections. The emphasis on detection increases the number of anomalies that are not targets and minimizes the likelihood that targets are missed.

Here we describe how the algorithms work best when targets are centered in the image frame and when there is little or no lateral platform motion. However, even in cases where these two conditions were not met, the adaptive algorithms detected targets reasonably well.

Background

As autonomous underwater vehicles (AUVs) become more ubiquitous in oceanic applications, so too do the requirements for improved sensing on these platforms. Forward-looking acoustic sensors are especially critical for object avoidance tasks. Acoustic sensing is an important attribute because optical systems tend to fail in turbid water; high-frequency acoustic systems are impeded much less by sediments suspended in the water. These systems are useful when tied to real-time processing schemes that can recognize objects to be avoided and cause the vehicle to be steered around them. In conjunction with onboard algorithms, the acoustic systems enable close-in inspection of objects; the vehicle makes decisions autonomously about objects being detected, and makes determinations of follow-on actions based on what has been found.

High-resolution, forward-looking acoustic sensors typically produce data that can be analyzed in image form. From these images, features about the object or scene being imaged can be estimated and used in classification or identification algorithms. A key feature type is the highlight caused by the reflection of the ping off the leading edge of the target. The bright return, especially when associated with an acoustic shadow directly

behind the highlight within the same beam, is an important indication that an object is proud (i.e., not buried in the seafloor).

The most important feature type for these kinds of sensors is the size and shape of the acoustic shadow cast behind an imaged object, formed because the object blocks the sound from propagating to the seabed while the surrounding bottom area scatters sound back to the receiver. Shadow size and shape, combined with information about the sensor position, can be used to estimate the size and shape of the imaged object.

Current Work

In section 2 of this report, we describe the sonar system used to collect the data sets examined here. In section 3 we discuss the development and analysis of the algorithms. A discussion of the data sets used and the results of our data analysis are given in section 4. The last section contains our conclusions about this work, and the Appendix gives our assessment of the requirements for future data collection.

2. DIDSON Description

The Dual-frequency IDentification SONar (DIDSON; Figure 1) is an emerging technology for high-resolution, forward-looking underwater sensing (*Belcher et al.* 2002a; *Belcher et al.,* 2002b; and the DIDSON website listed in **References**). DIDSON is an acoustic lens sonar system, forming a set of effective horizontal beams by steering sound coming from particular directions via a set of acoustic lenses onto a single acoustic receiving element for each effective beam. This eliminates the need for electronic beam forming, and allows the DIDSON to consume only 30 W of power. The two operating frequencies are 1.0 and 1.8 MHz. DIDSON has a total horizontal field of view of 29° at both operating frequencies. The two-way horizontal beam widths are 0.4° and 0.3° at 1.0 and 1.8 MHz, respectively. The horizontal beam spacing is 0.6° at 1.0 MHz and 0.3° at 1.8 MHz. The two-way vertical beam width is 12° at both operating frequencies. The DIDSON outputs 48 and 96 effective beams of data at 1.0 and 1.8 MHz, respectively. Using wideband transmit pulses, down-range resolutions of approximately 3.5 cm at 1.0 MHz and 1.8 cm at 1.8 MHz are realized. Each beam consists of 512 range bins in a raw image, so a beam space image consists of 96 x 512 pixels in the high-frequency mode, and 48 x 512 pixels in the low-frequency mode.

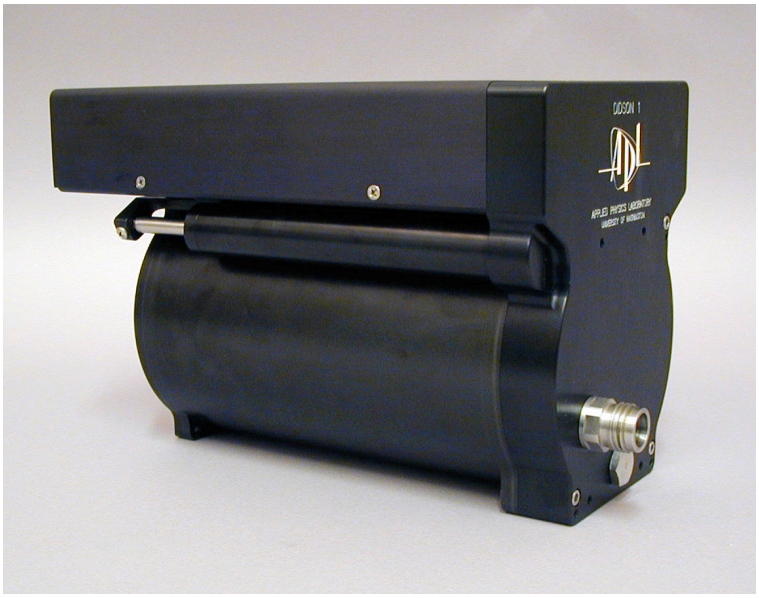

Figure 1. The Dual-frequency IDentification SONar (DIDSON)

Each data frame of DIDSON imagery is formed using multiple pings of the sonar. For high-frequency images there are eight pings with 12 beams in each ping. Low-

frequency images are generated from four pings also of 12 beams each. The pings and the beams within each ping are spaced and ordered to minimize the cross talk among the beams, a common problem when generating images from multiple beams. When the sonar platform is moving, each ping occurs at a slightly different location and the resulting beams of data do not align with the beams in the previous ping. Forward motion causes a beam's subsequent pings to begin at a different range. Changes in platform heading can cause multiple pings to cover the same ground and to miss stripe-shaped regions on the seabed. Heading rate changes to the left generally decrease beam spacing and result in a smaller field of view with a possible gap in the center of the frame. Heading rate changes to the right generally increase beam spacing and result in a larger field of view with overlapping beams at the center of the frame and gaps towards the sides. Large heading rate changes can even reorder the beams.

DIDSON has a "sweet spot" in the center of its field of view approximately one-third of the total frame length from the start of the frame and continuing until approximately one-quarter of the total frame length from the end of the frame (although these values are somewhat dependent upon the sonar tilt and range of the region of interest). Outside this sweet spot, attenuation from transmission loss (TL) and the vertical beam pattern of the acoustic lens reduces the signal-to-noise ratio to an unproductive degree. Ideally for classification and identification, the target and its complete shadow should occupy the sweet spot of the frame.

Figure 2 illustrates a raw DIDSON image of a rock sitting proud on a flat, sandy sea bottom. Note that the target is centered in the field of view. Dark areas on the edges are due to uneven azimuthal illumination. The jagged lines are caused by forward motion of the sonar platform, and the dark foreground can be attributed to a sharp cut off of the vertical beam pattern. Section 3 discusses preprocessing steps that correct for some of these effects and discusses follow-on image and information processing algorithms.

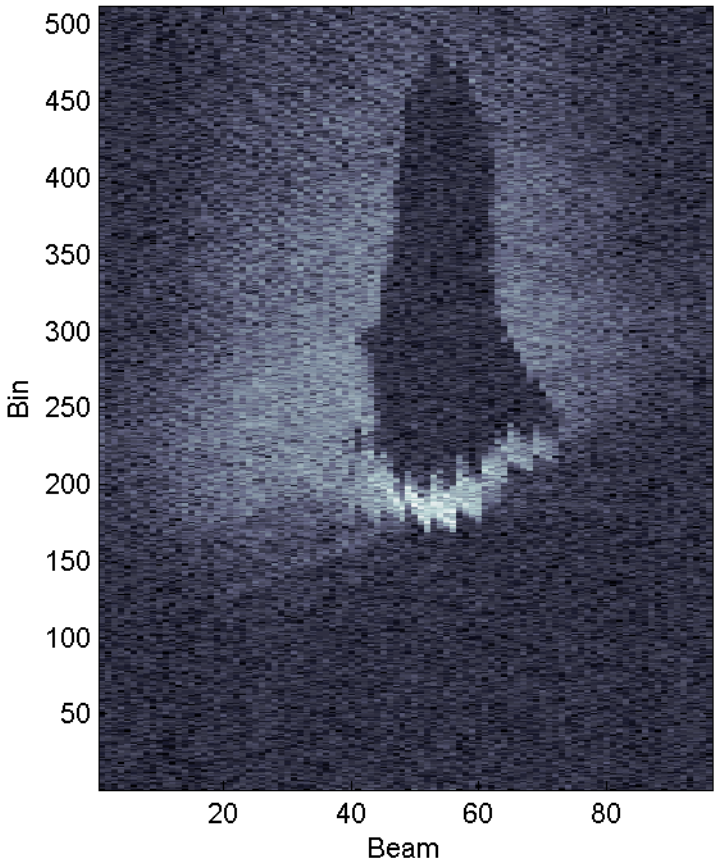

Figure 2. Raw DIDSON image of a rock sitting proud on the seabed

3. Development and Analysis

Normalization

All images from the data required initial preprocessing to make them consistent, which in turn improves the likelihood of successful detection, highlight localization, and shadow segmentation. The initial preprocessing consists of normalization and motion correction. Normalization, which equalizes illumination characteristics, varies depending upon the processing chain for either highlights or shadows, but motion correction, which compensates for any motion of the sonar platform, i.e., AUV, is common to both processing chains (Figure 3).

Normalization is the process that attempts to remove broad underlying trends that have no information content for follow-on image and information processing from the image data before feature extraction processing can occur. Normalization procedures must estimate the local background variation accurately, so that when this background variation is removed, the salient features of the image can be extracted.

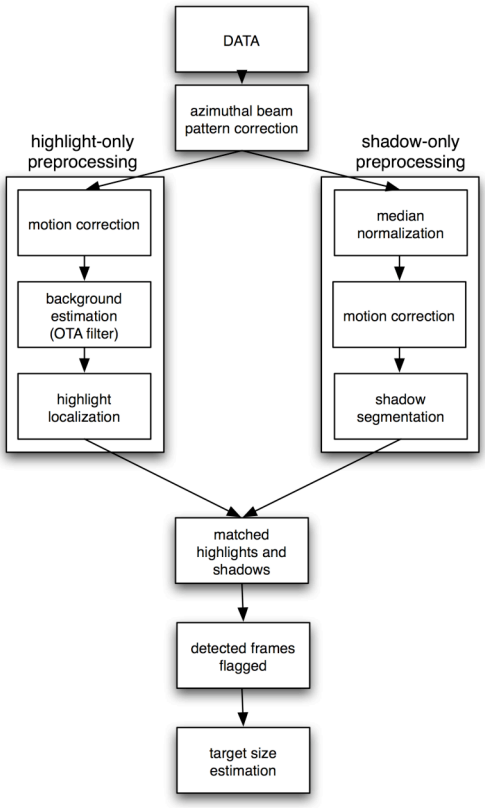

Figure 3. Signal, image, and information processing chain for target detection

We corrected for two factors that contributed to the background variation. The first was beam-to-beam variation wherein we used an azimuthal beam pattern correction that was based on measured beam pattern data. The azimuthal beam pattern correction fixes the non-linear drop-off in intensity at the sides of the image that are inherent in the acoustic lens of the sonar. This was not a perfect correction, especially as ambient noise became a major contributor to the received energy level, but this correction removed a significant portion of the azimuthal variability. Figure 4 shows the results of azimuthal beam pattern correction to the rock image (Figure 2).

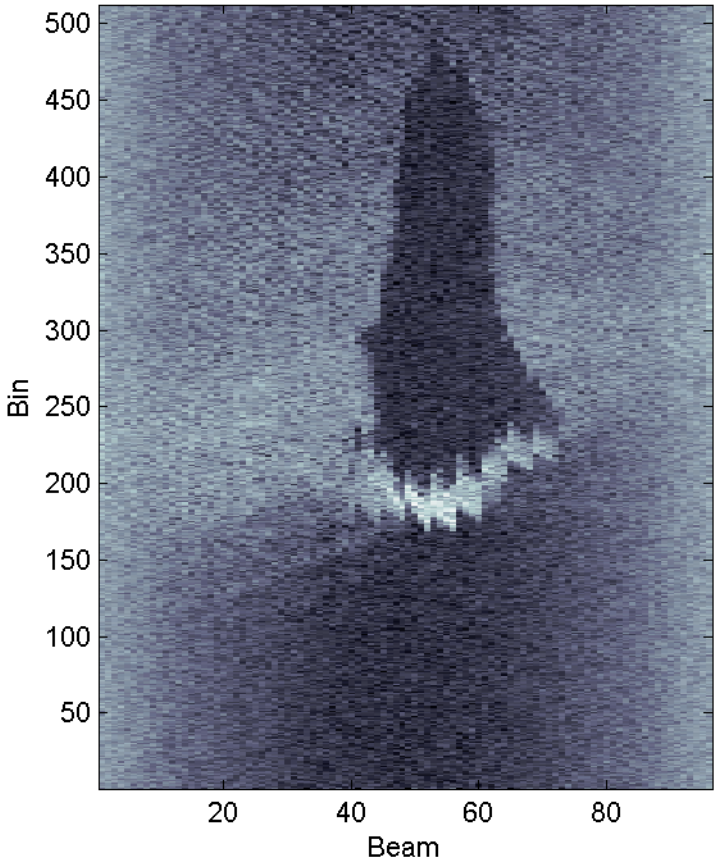

Figure 4. Azimuthal beam pattern correction for rock image (Figure 2)

The second factor that contributed to background variation was within-beam variation caused by the vertical beam pattern effect and transmission loss (TL). The vertical beam pattern effect causes a lower intensity foreground where the sound is not yet reflected from the bottom. The attenuation due to TL causes the intensity to drop off

as distance from the sonar increases. Within-beam background variation needed to be corrected or normalized differently depending upon whether we were processing for highlights or shadows. The background estimate used in highlight preprocessing eliminates the need to separately correct for attenuation and TL.

Normalization in preparation for shadow segmentation must be performed carefully. The intent is to preserve shadow-to-reverberation contrast, so that shadow segmentation may be calculated robustly in the following procedures. Because shadow regions may occupy a significant portion of a DIDSON image, these normalization procedures become non-trivial. We developed a normalization algorithm that worked well to preprocess DIDSON images for shadow segmentation (*Fox et al.*, 2004).

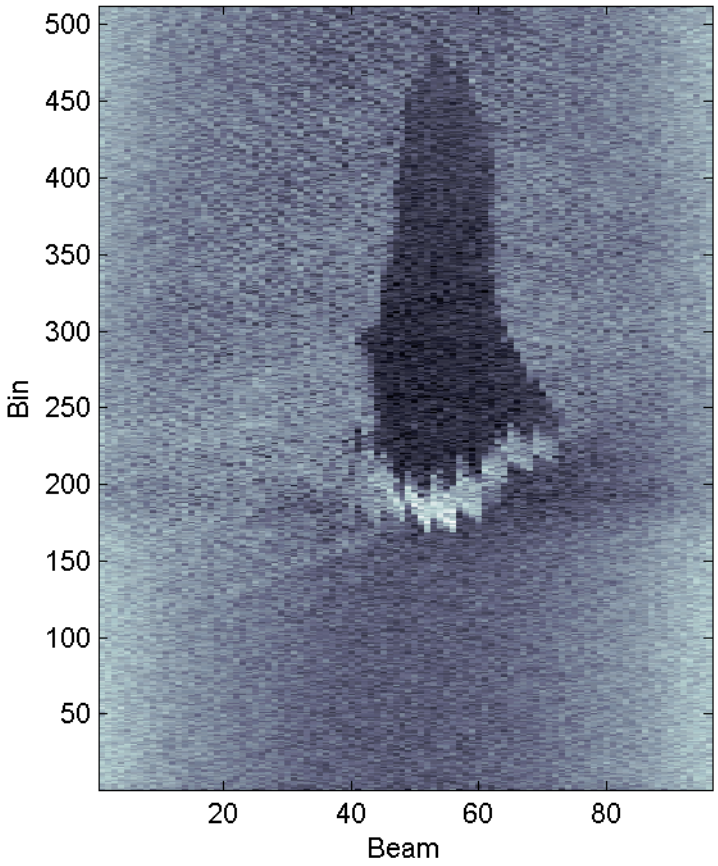

Figure 5. Shadow normalization of rock image (Figure 2)

First, we made an adaptive estimate of the range-varying portion of the received energy. At each range bin the background level was estimated by calculating the median level across beams. Next, we smoothed the series of median values to produce a single range-varying power profile that was removed from each beam to produce the normalized image (Figure 5). We found this order statistic approach to be very important for maintaining shadow-to-reverberation contrast in DIDSON images.

Motion Correction

We compensated for the motion of the sonar platform because each image was not obtained instantaneously, and, when the sonar platform moved, our assumptions also changed as to where each pixel of data belonged within the image. Note the effects of platform motion in Figures 4 and 5, which are manifested as jagged lines and reduced resolution.

Motion correction consists of correcting for the forward motion of the platform and for heading rate change. To compensate for the forward velocity of the sonar, we used the estimated platform speed to map pixel intensity into a registered range bin. Each beam was processed independently. When there was a change in the heading rate, the cross-range resolution decreased. The correction was dependent upon the direction of the heading rate change, which could change the beam ordering from the DIDSON. Using knowledge of the timing of the eight pings and the heading rate change, we remapped the beam order and their relative locations. Finally, we interpolated the data bilinearly to obtain equally spaced beams. A significant change in heading can cause a reduced field of view, resulting in fewer than 96 beams of data available for processing.

Figures 6 and 7 illustrate the results of motion correction on the azimuthal beam pattern corrected data (in preparation for highlight normalization) and on the shadow normalized data.

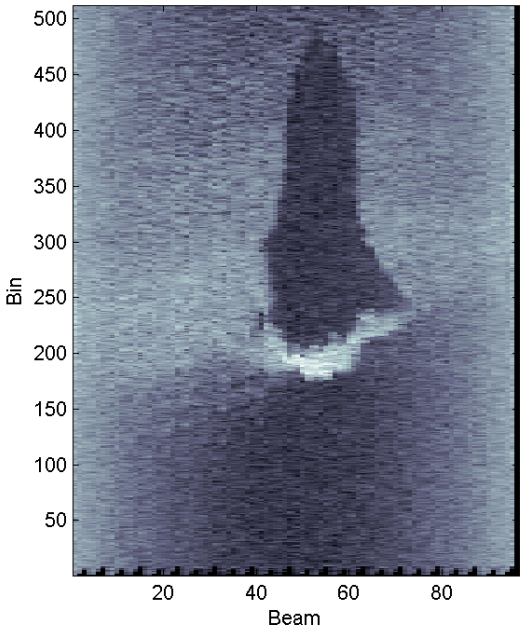

Figure 6. Azimuthal beam pattern and motion corrected rock image (Figure 2)

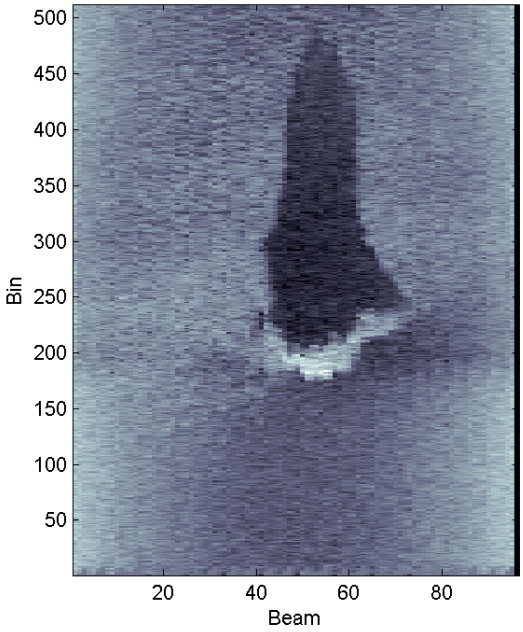

Figure 7. Shadow normalized, motion corrected rock image (Figure 2)

The motion correction algorithm is dependent upon the accuracy of the velocity estimate. We found the sonar platform motion estimates limiting. One experiment used the cross-correlation of adjacent frames to estimate the forward velocity of the sonar platform. We found that when there were strong bottom features, the cross-correlation resulted in a more accurate estimate of velocity than available with the vehicle sensors, but when the seabed lacked strong features, the estimates were significantly worse than the sensor data. Overall, the significant increase in processing time and restricted usefulness dissuaded us from using the cross-correlation estimate of forward velocity in the present processing string.

Background Estimation

After azimuthal beam pattern and motion correction, highlight preprocessing consisted of performing background estimation, background removal, and thresholding to localize or find the highlights. The background estimate was calculated using an order truncate average (OTA) filter (*Struzinski and Lowe*, 1984). We found that three of the four background normalization schemes Struzinski and Lowe proposed for signal detection systems showed promise. The three we tested were: two-pass mean (TPM), split-average exclude-average (SAXA), and OTA. We found that the OTA filter, when confronted with a wide peak, was the least likely to become biased, i.e., incorporate any of the signal into the background estimate. This normalizer was the most processor intensive, but OTA was selected because its accuracy more than compensated for a slight increase in processing time compared to SAXA. The ability of TPM to track the background was inadequate.

Our implementation of the OTA filter used a window of 51 bins. For each bin in a beam, the sample median of the window was calculated. Then, the sample median was multiplied by a shearing threshold. Lastly, all values greater than the threshold were excluded from a noise mean estimate, which was calculated as the mean of the remaining values. Because of the window size, the first and last 25 bins of each frame were discarded. We ran each beam of the image independently.

Figure 8 illustrates the highlight normalized image of the rock shown in Figure 2. Note that the shadow in this image is barely discernible compared to the shadow in the shadow normalized image (Figure 7; i.e., shadow-to-reverberation contrast is maintained by the shadow normalization, but not the highlight normalization).

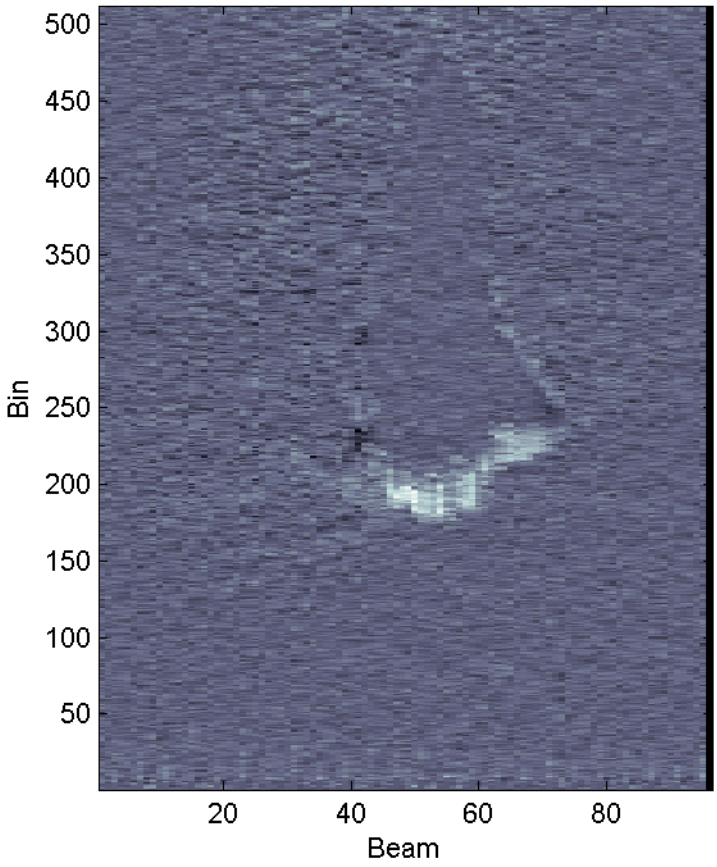

Figure 8. Highlight normalization of rock image (Figure 2)

Highlight Localization

Using the background normalization estimate, we defined highlight detection as any return that was 17 dB greater than the background. This value was calculated from testing the values between 13 dB and 21 dB using several data sets (described in section 4). This threshold value provided the best tradeoff between highlight localization and false alarms. We developed a clustering filter to remove isolated noise pixels while keeping potential highlight localization. The filter checks every 5 x 5 pixel region. If at least eight pixels in this region were highlights, all the highlights in the region were copied. If less than eight pixels in a region were highlights, then none of the highlights were copied at that time. If the highlights belonged to a cluster, they were copied when the highlights were centered in a region. The highlight localization pixels were grouped

using connected components. Then each group's location and extent were calculated to link to shadows in matched highlights and shadows. Figure 9 illustrates the effects of highlight localization on the rock image from Figure 2.

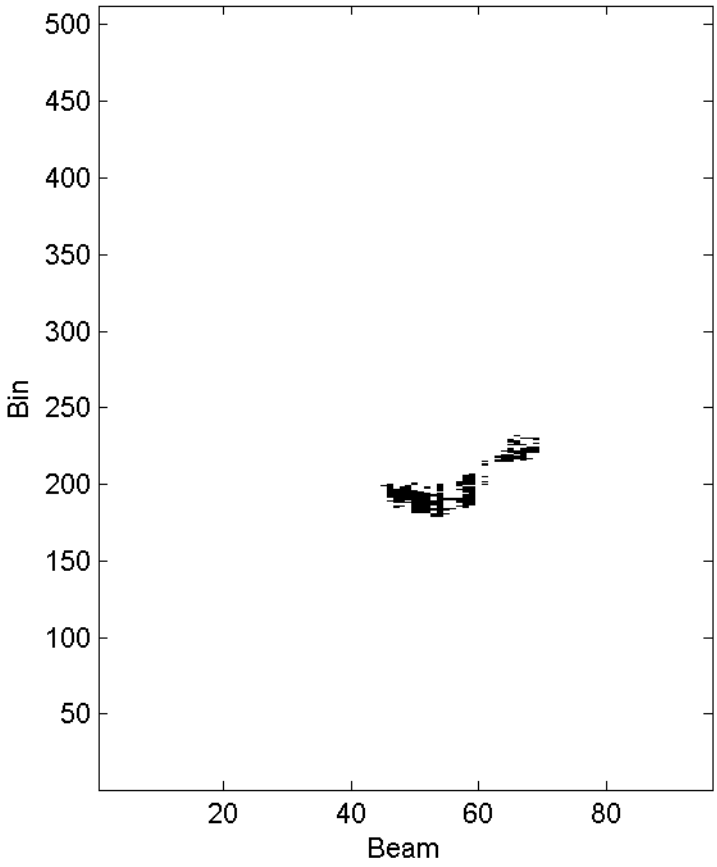

Figure 9. Localized highlights from rock image (Figure 2)

Shadow Segmentation

Differentiating an object's acoustic shadow region from normal reverberation background energy is one of the more challenging image processing tasks. Thresholding is a standard technique used in image processing that segments an image by setting all pixels whose intensity values are above a threshold to a foreground value and all the remaining pixels to a background value. The conventional method utilizes a global threshold for all pixels. The varying background in our images from uneven illumination

and seafloor texture made the task all the more difficult because a single threshold value represented shadow in one region of the image, yet background in another.

Adaptive thresholding changes the threshold dynamically over the image assigning each pixel its own threshold. Dynamic thresholding uses a lower threshold in dark areas of the image and a higher threshold in light areas of the image. In most cases, dynamic thresholding does a good job of preserving high-frequency details such as edges, even in images of subjects with uneven illumination where edges may move through both light and dark areas of the image. (For further information please see the adaptive and dynamic websites listed in **References**.)

We adapted a dynamic thresholding technique from biomedical imaging developed by *Chow and Kaneko* (1972) to segment acoustic shadows. The algorithm is dynamic in that different regions of the image are analyzed independently (although regions overlap and smoothing processes are used in the algorithm), and thresholds are set based on the local statistics of the image.

The algorithm begins by dividing an image into a set of smaller overlapping regions of equal size. *Brodsky et al*. (2004) reports that a major problem is in determining the number of thresholded subregions, or the proper region size. When considering mine-like objects in DIDSON images, the size of these regions affects the algorithm's sensitivity to local variations in the signal return strength and the ability to delineate the actual shadow. Here region size was a compromise between the need to reflect a density and diversity of values while limiting the impact to a local area. Each image was divided into 25 by five overlapping regions of approximately 32 beams by 39 range bins. This region size provided enough pixels to generate a representative histogram and localized the resulting threshold to minimize noise.

Subsequently, we computed a histogram for each region. When boundaries between shadows and background were present in a particular region, the histogram was expected to be bimodal. The histogram of each region was therefore modeled as a sum of two Gaussian distributions, and the parameters of the Gaussian sum (means, variances, and coefficients of mixture) were estimated with an expectation maximization (EM) algorithm. Note that these distributions are assumed Gaussian in dB space. Although there is no good physical justification for this assumption in the case of acoustic reverberation and shadow formation, this was the formulation used in the original work, and the results here were adequate. See *Fox and Hsieh* (2005) for further discussion and analysis of reverberation statistical distributions in DIDSON data.

Each region's histogram was checked for bimodality by its valley-to-peak ratio and its mean difference. The valley-to-peak ratio is a measure of the dip between the two

peaks, and is used to measure the separateness of the two distributions. The mean difference is a calculation of the distance between the two peaks, and helps ensure that the two peaks are distinct. Thus, for regions with bimodal distributions a threshold can be calculated by solving a quadratic equation derived from the method of maximum likelihood, which allows us to minimize the probability of misclassification.

For regions with only one distribution, no threshold was calculated because all the pixels in that region were assumed to belong to the same class. Their threshold values were assigned by a weighted average of the computed thresholds from their neighboring regions. A threshold value was specified to ensure that a minimum number of neighbors with computed thresholds were used to calculate an averaged threshold. This also yielded an averaged threshold, which smoothed the results, to be used for regions with a computed threshold.

Once we established a threshold for each region we further smoothed the images by assigning the threshold to the centroid of each region. The remaining pixels' threshold values were calculated with bilinear point-wise interpolation.

After the dynamic thresholding process, some nuisance pixels typically still remain. Speckling from multiplicative noise and bottom texture can contribute to noise in the segmentation. The segmentation is cleaned up with a morphological opening process (*Pratt*, 1991) that employs a 3 x 3 structuring element to remove isolated pixels and smooth the remaining larger features.

Figure 10 illustrates the segmentation resulting from our dynamic thresholding algorithm. Note there is an apparent shadow in the foreground as well as behind the highlights shown in Figure 8.

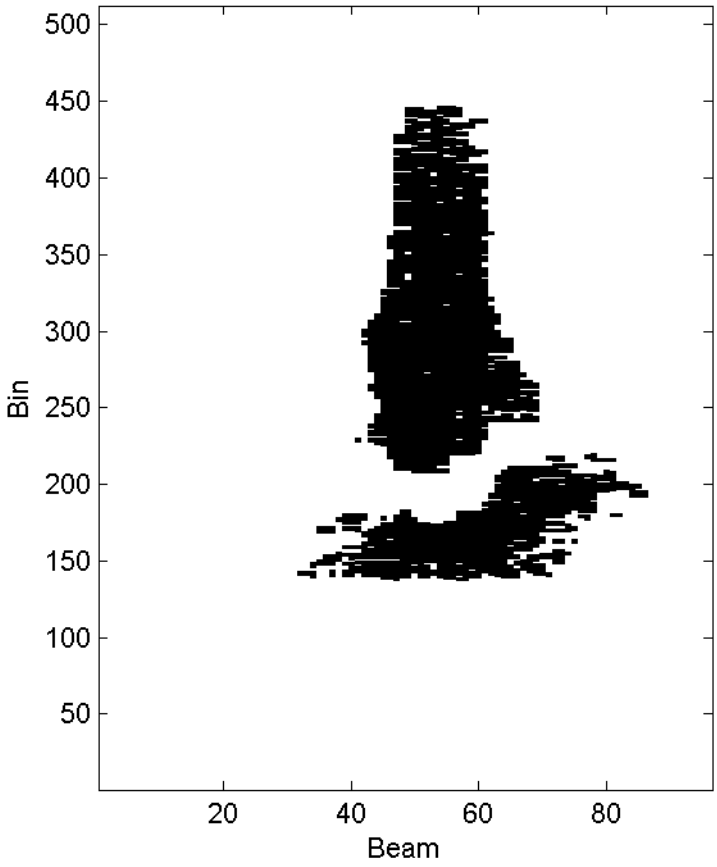

Figure 10. Shadow segmentation of rock image (Figure 2)

To further remove spurious structures, we used our knowledge of possible target shadow size to eliminate connected regions with an area smaller than 250 pixels. Then we calculated the width, length, and extents, i.e., starting and ending bin, first and last beam, of the remaining segmented shadow regions. At this point we had the information necessary to match highlights to shadows, flag the detected frames, and estimate target size.

Matched Highlights and Shadows

Typically, the target reflecting the ping of the sonar was visible in the image as a group of highlights. The decreased return following the reflection was viewed as a shadow. Along a beam the peak of a highlight, followed by a corresponding valley of an

acoustic shadow, could be used, theoretically, to detect targets. We found this single-beam approach to be non-robust. One way to distinguish a shadow from background, taking into account the variability of the background, is to use information from adjacent beams that have a similar background. However, it was not uncommon for a target shadow to consist of a significant portion of the beams available in a given frame, making it difficult to use the adjacent beams to determine a background estimate. Therefore, instead of detecting targets on a beam-by-beam basis, we searched the region behind each detected highlight for a shadow segmented by dynamic thresholding. When a sufficient overlap was found between the beams of the highlight region and the beams of the shadow region, detection was triggered.

4. Data Sets and Results

Data

The data sets were collected with a DIDSON mounted in a forward-looking configuration on a REMUS autonomous underwater vehicle (AUV) (*Von Alt et al.*, 2001). Much of the data analyzed in this report are from runs over a relatively flat, slightly rippled seabed, with the vehicle's forward speed set at approximately 2.0 m/s, and its height above the seabed set at 2–3 m. The downward tilt of the sonar was approximately 17°.

The primary data set used was from the Battlespace Preparation 2002 (BP02) exercise, collected off the coast of Elba, Italy. We also used data from Fleet Battle Experiment Juliet (FBE-J) in 2002, collected off the coast at Camp Pendleton, CA. These data sets were selected for their navigation system data and exercise mine shape targets. The only data set collected by an AUV with truth (certainty of the target types) is the AUV Fest 2003 data set, which contains mine-like objects but no exercise mines. Other data sets available, but not used, showed exercise mine targets, but were diver-held and lacked data in the platform status fields for latitude, longitude, depth, altitude, velocity, heading, heading rate, pitch, and roll. It should be noted here that the lack of truth and low numbers of targets somewhat limited the scope of this work.

A target was labeled as such during a visual inspection of the data if an object looked to our eyes (untrained in target detection or identification) like something an operator would probably want a second look at. We examined the frames before and after the probable target and marked the first and last frame that could be identified as something of interest. A target can traverse multiple frames, and in fact, all our targets did.

We did not attempt development of an algorithm for low-frequency DIDSON data. The low-frequency data proved to be sufficiently different, such that a separate set of tuning parameters needed to be estimated for the algorithms. There were too few low-frequency targets in the data available to do this estimation and proper algorithm testing.

The BP02 data set contains 7,660 frames of 1.8-MHz data. After manual truthing of these frames, we concluded there were 16 potential targets in 126 frames of high-frequency data. The bottom types varied among sandy bottom, distinct sand ripples, some kind of biomass, and scattered large rocks.

The FBE-J data set contains 33,169 frames of 1.8-MHz data. Manual truthing revealed 14 potential targets in 126 frames of high-frequency data. The bottom type in this data set was mostly sandy bottom, some flat and some rippled.

The AUV Fest data set contains 30,260 frames of 1.8-MHz data over a generally flat sandy bottom with significant numbers of what appear to be fish in some images. After manual truthing of these frames, we concluded there were 33 targets in 250 frames of high-frequency data.

Results of Data Analysis

Our algorithms used on the BP02 data set detected 14 out of 16 targets. The false alarm rate, i.e., the percentage of frames not labeled as targets that triggered a detection by the algorithm, was 3.9%. The two targets that were not detected by the algorithm did not have a shadow that was segmented. Some of the false alarms occurred when rocks or other clutter on the seafloor had shadows that were similar in size to what we expected to be mine-like objects. Figures 11–14 show an undetected target with a shadow that could not be segmented as the frame proceeds through the processing chain. Although this target appeared in six frames, it is only visible in the upper left corner, a difficult location for the automatic algorithms to process.

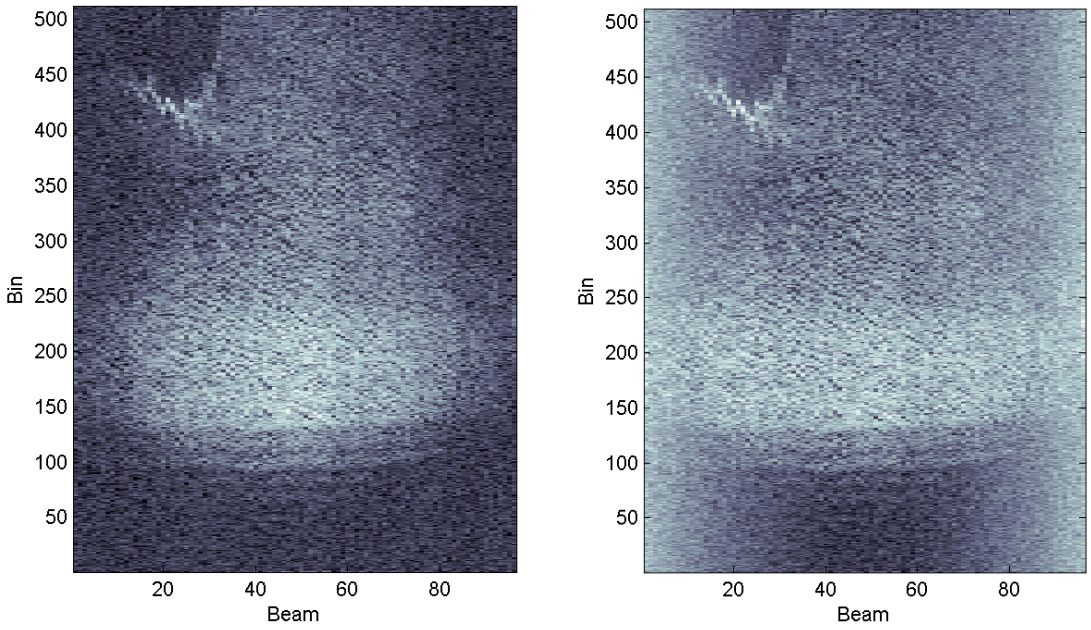

Figure 11. (right) Raw image of undetected target and (left) azmiuthal beam pattern correction applied to image

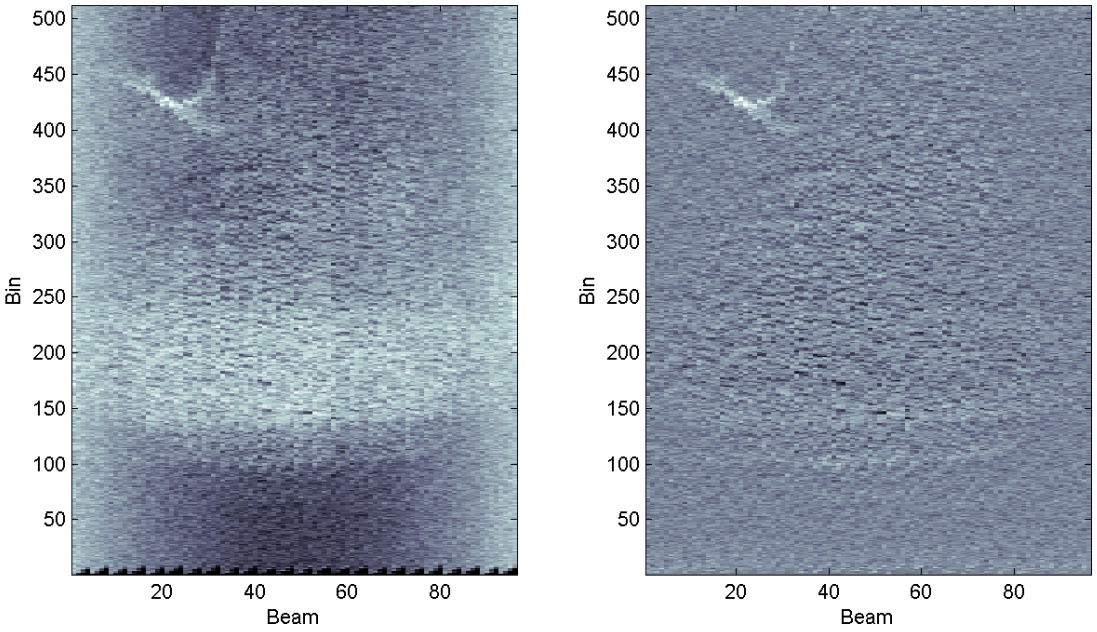

Figure 12. (right) Azimuthal beam pattern and motion corrected image of undetected target and (left) highlight normalized image of undetected target

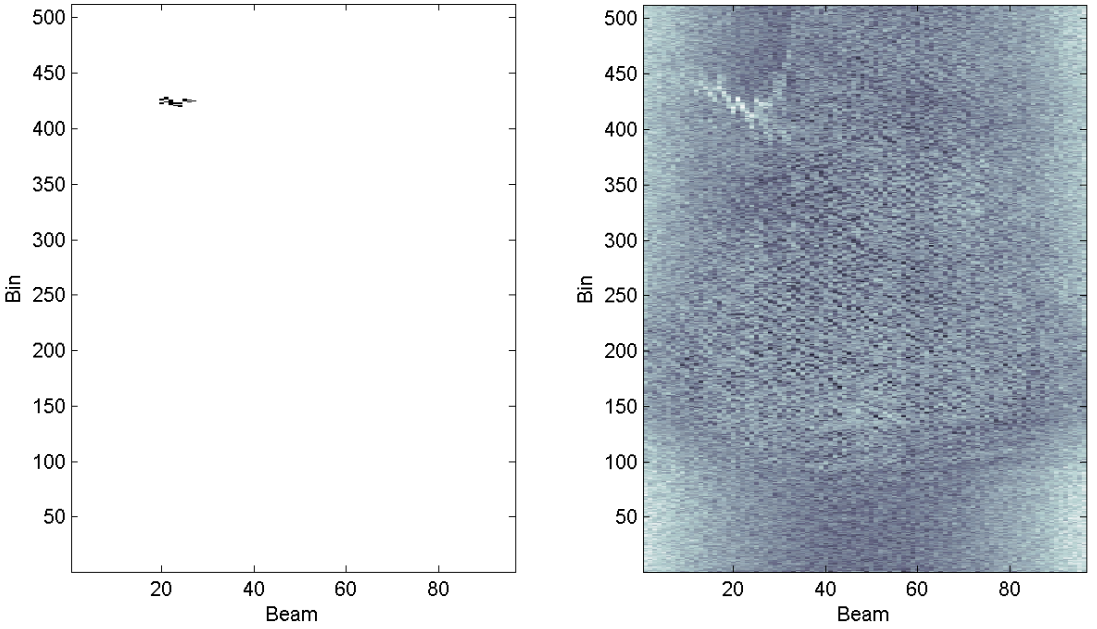

Figure 13. (left) Highlight localization of undetected target and (right) shadow normalized image of undetected target

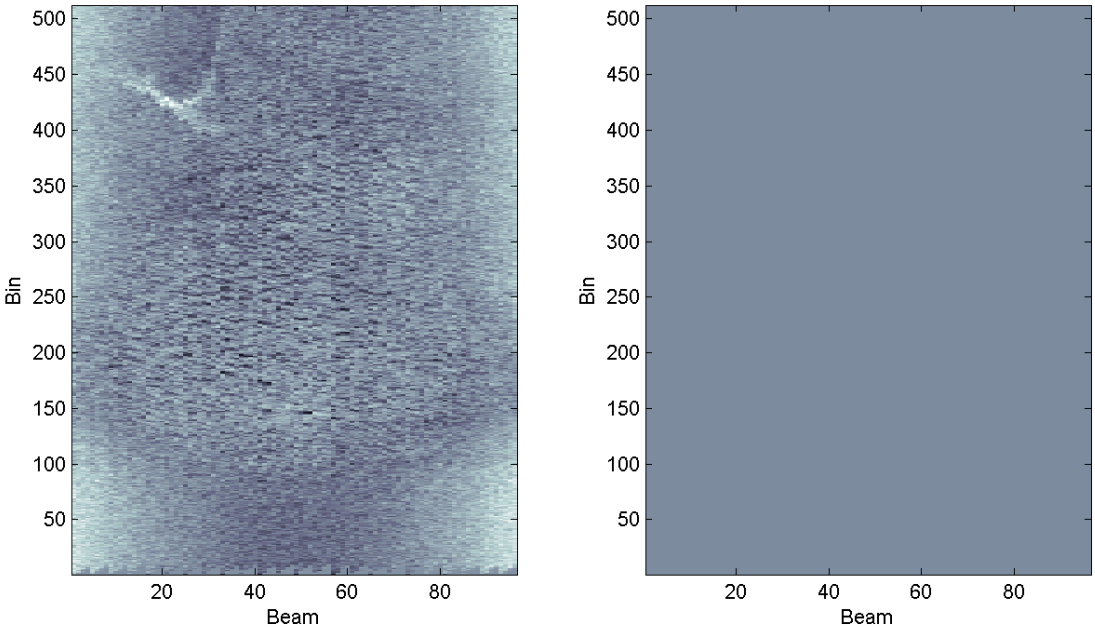

Figure 14. (left) Shadow normalized, motion corrected image of undetected target and (right) undetected target with shadow that could not be segmented

Figures 15–18 illustrate a false alarm that exhibits highlight (biomass, perhaps fish) at the center of the image followed by a shadow caused by bottom texture. An experienced operator may or may not agree with our assessment.

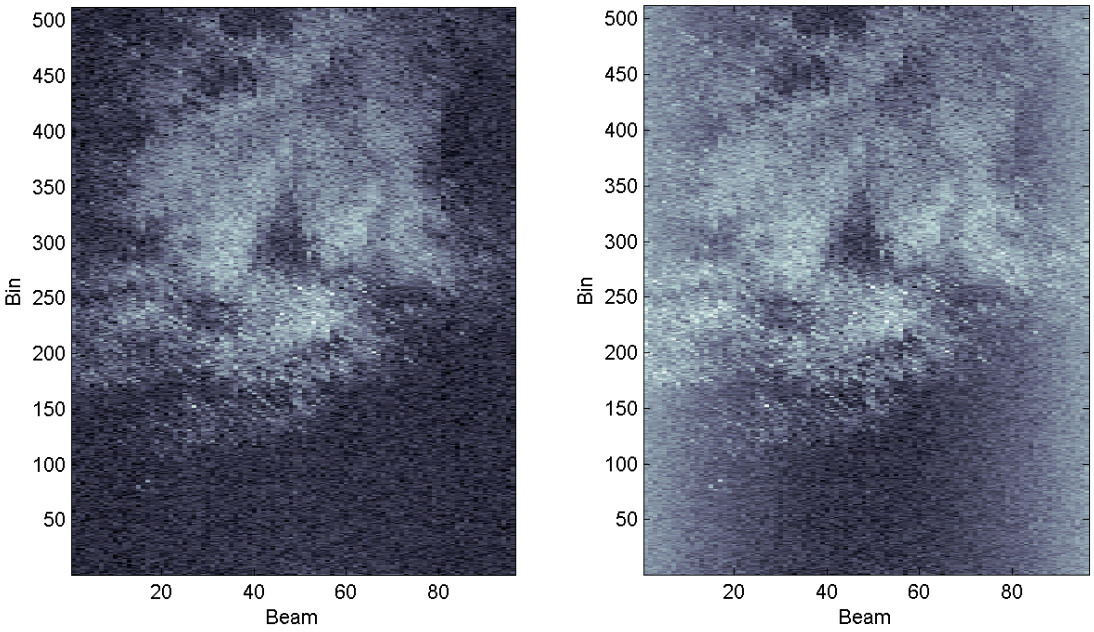

Figure 15. (right) Raw image of false alarm and (left) azimuthal beam pattern corrected image of false alarm

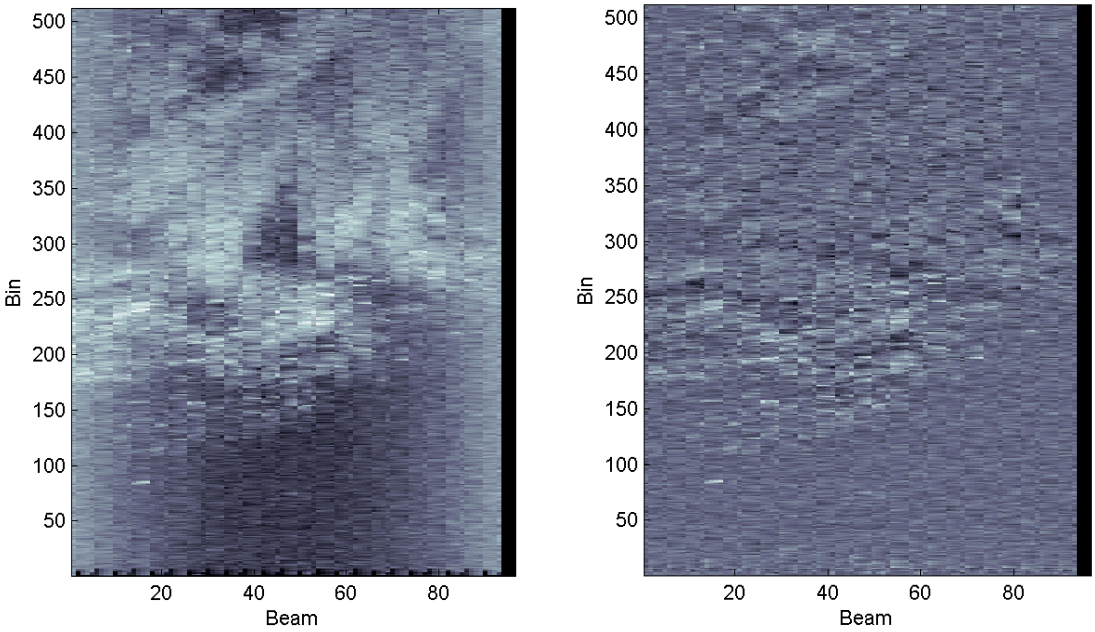

Figure 16. (right) Azimuthal beam pattern and motion corrected image of false alarm and (right) highlight normalized image of false alarm

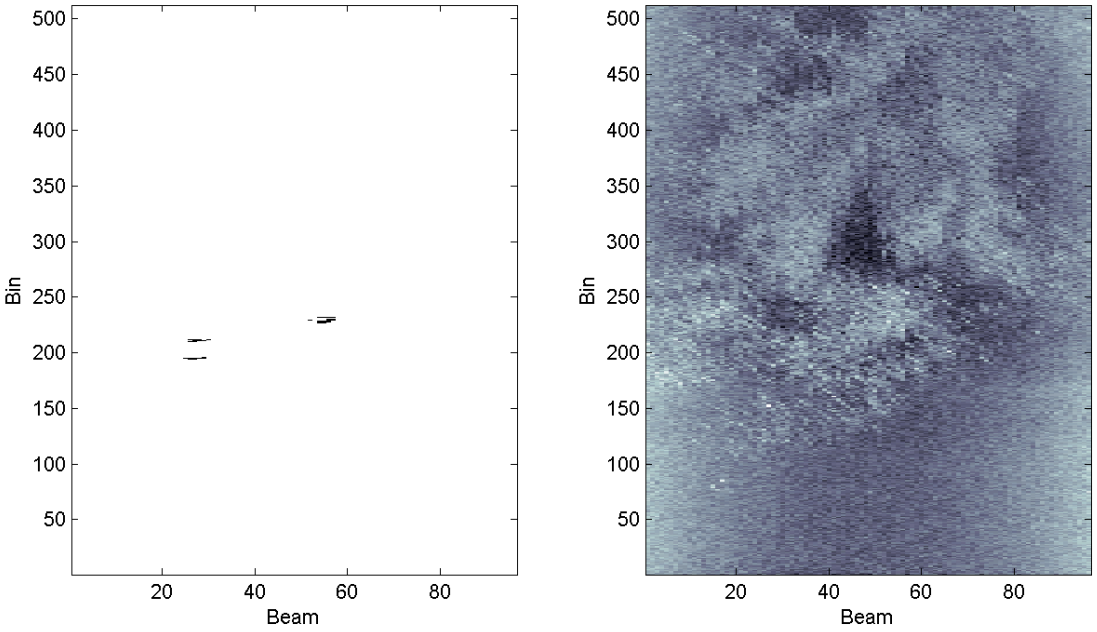

Figure 17. (left) Highlight localization of false alarm image and (right) shadow normalization of false alarm

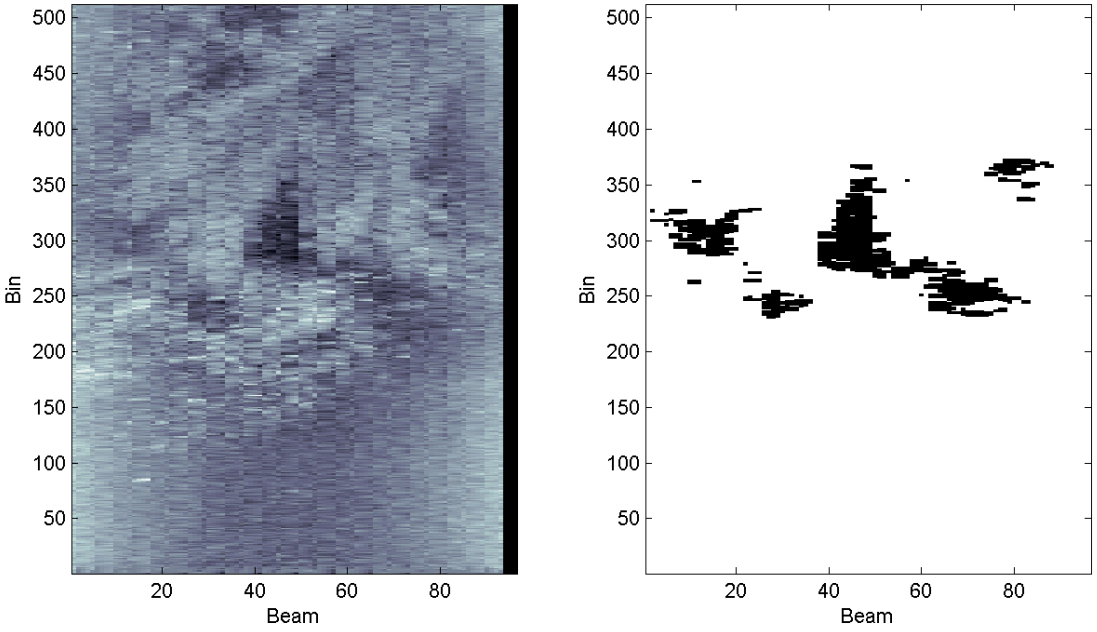

Figure 18. (left) Shadow normalized, motion corrected image of false alarm and (right) shadow detected image of false alarm

Ripples on the seafloor also triggered our detection algorithm. Over 60% of the false alarms in the Italy data set were triggered by ripples on the seabed. At the writing of this report we are applying heuristic techniques to the results in hopes of reducing the number of false alarms from ripples. These methods include a comparison of the highlight extent with the shadow extent. Results are inconclusive at this time. The algorithms require further refinement.

Processing the FBE-J data set resulted in 12 out of 14 detections with a false alarm rate of 1.6%. Note that the FBE-J experiment site had a mostly flat sandy bottom with less clutter, and a corresponding lower false detection rate. Again, the algorithm failed to segment a shadow for the two targets that were not detected.

From a total of 30,260 high-frequency frames, our results for the AUV Fest 2003 data set show 18 of 33 target detections with a false alarm rate of 0.3%. One of the missing targets appeared to be a wire mesh crab trap that did not image like any of our other targets: line-like bright highlights, but no solid shadow. Most of the remaining targets that were not detected appeared briefly in the far corners with weak shadows that were difficult to segment. The sonar window start and length settings also may have contributed to the difficulty the algorithm had segmenting shadows on this data set. The other data sets typically had a window start setting of 3.75 m or occasionally 5.63 m and a window length of 9 m. The AUV Fest 2003 used the window start settings of 1.88 m for most of the data and 5.63 m for the remainder.

Target Size Estimation

Once a detection was made, an estimate of the target size could be calculated with the width and length of the segmented shadow. The width of the shadow should be representative of the width of the object. For object height estimation, the data must first be transformed into Cartesian space, and the true shadow length (down range) along the seafloor determined. Figure 19 illustrates that from the geometry, we can estimate the height of the object using the shadow length, location, and the elevation of the REMUS vehicle above the seafloor.

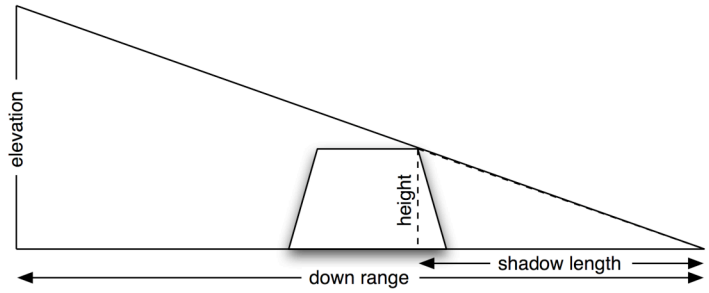

Figure 19. Geometric relationship of the sonar to the object

The precision of the width estimate was limited by the cross-range resolution of the sonar data. Using the high-frequency identification mode of the DIDSON (1.8 MHz), the cross-range resolution at a distance of 10 m from the sonar is approximately 5.2 cm.

The certainty of the height estimate was more complicated. It is dependent upon the accuracy of the navigation system measurements, primarily the elevation above the seafloor that was used in the calculation of the object height as well as in the conversion to the true shadow length along the seafloor. The velocity of the sonar platform also may affect the quality of segmentation and the remapping of the bin location during motion correction.

Height estimates from three sample frames were 0.48 m, 0.46 m, and 0.49 m (*Fox et al*. 2004). These estimates are reasonably consistent, although biased somewhat low compared to the true object height (~0.51 m).

5. Conclusions

Numerous promising results emerged, but the available data sets limited the scope of this project; all data analyzed here were collected with a single DIDSON mounted on a REMUS. The small number of data sets limits our judgment of how well the algorithms worked in a variety of operating conditions, and the relatively few and unidentified targets made it difficult to develop the algorithms.

Nevertheless, what we were able to determine was that the quality of the data is sensitive to motion. Accurate estimates of velocity are required to minimize data uncertainty. If possible, heading rate change should be kept to a minimum because it can reduce the cross-range data and resolution. Also preferable are slower speeds and/or higher frame rates when a target is in view. A concept of operations that allows for hovering and slowly circling a target upon reacquisition may take better advantage of the features of the DIDSON. (A more complete explanation can be found in the Appendix.)

The DIDSON's sweet spot, centered in cross-range and between one-third of the front edge to one-quarter of the back edge of the image, has the greatest signal-to-noise ratio. Detection was difficult when the targets were not centered in the frame. Many of the missed detections were due to targets that crossed the field of view only in the far corners where shadow segmentation was especially difficult. Despite all this, the adaptive algorithms were effective in target detection.

In section 3 we describe a shadow segmentation algorithm that makes very specific assumptions about the form of the reverberation (and shadow) envelope statistics. We do not have statistical confirmation that our Gaussian assumptions are true; they must be verified in the future. We anticipate that a more complete understanding of the reverberation statistics in this frequency band will enable enhancements to the algorithms that can improve performance.

6. References

Adaptive Thresholding website: http://www.cee.hw.ac.uk/hipr/html/adpthrsh.html

Belcher, E.O., W.L.J. Fox, and W.H. Hanot, "Dual-frequency acoustic camera: a candidate for an obstacle avoidance, gap-filler, and identification sensor for untethered underwater vehicles," *Proceedings, Oceans '02,* ***4***, 2124–2128 (MTS/IEEE, 2002a).

Belcher, E.O., W. Hanot, and J. Burch, "Dual-frequency IDentification SONar (DIDSON)," *Proceedings, 2002 International Symposium on Underwater Technology*, 187–192 (2002b).

Brodsky, P.M., J.C. Luby, and J.I. Olsonbaker, *Innovative 3D Visualization of Electro-optic Data for Mine Countermeasures*, APL-UW Technical Report 0401, March 2004, pp. 38–39.

Chow, C.K., and T. Kaneko, "Automatic boundary detection of the left ventricle from cineangiograms," *Computers and Biomedical Research,* ***5***, 388–410, 1972.

DIDSON website: http://www.soundmetrics.com/.

Dynamic Thresholding website: http://imagebeat.com/thresh.html.

Fox, W.L.J., J.B. Hsieh, and C. Polwarth, "Segmentation of images from an acoustic lens sonar," *Proceedings, Oceans '04, MTS/IEEE Techno-Ocean '04*, ***4***, 2029–2035 (MTS/IEEE, 2004).

Fox, W.L.J., and J.B. Hsieh, "Reverberation statistics at 1.0 and 1.8 MHz measured by an acoustic lens imaging sonar mounted on an AUV," *Proceedings, 1st International Conference on Underwater Acoustic Measurements: Technologies and Results*, *Heraklion, Crete, June 2005*.

Pratt, W.K. *Digital Image Processing*, 2nd ed. New York: John Wiley & Sons, Inc., 1991, pp. 449–472.

Struzinski, W.A., and E.D. Lowe, "A performance comparison of four noise background normalization schemes proposed for signal detection systems," *Journal of the Acoustical Society of America*, ***76***, 1738–1742, 1984.

Von Alt, C., B. Allen, T. Austin, N. Forrester, R. Goldsborough, M. Purcell, and R. Stokey, "Hunting for mines with REMUS: A high performance, affordable, free swimming underwater robot," *Proceedings, Oceans 2001*, ***1***, 117–122 (MTS/IEEE, 2001).

7. Appendix

In the process of developing our target detection algorithms while working with DIDSON data, we have arrived at several recommendations about data collection requirements needed to support autonomous identification.

To obtain the detailed imagery necessary for identification purposes, the concept of operations used to gather the existing data sets requires revision. Sonar platform motion significantly degrades the quality of DIDSON images. Data obtained by divers have offered a tantalizing view of imagery that might be possible with a hovering vehicle. To best utilize the DIDSON imaging sonar, it must be integrated with a platform that has the ability to hover near a target.

DIDSON's high-frequency mode at 1.8 MHz provides the necessary resolution for identification. The low-frequency mode at 1.0 MHz does not provide enough detail, but can be useful to reacquire the target. Because sonar illuminates one side of the object and the other side remains obscured in shadow, multiple aspects will be necessary to utilize all available information about the target.

Changes in heading rate decrease the cross-range resolution because of the multi-ping image generation used by the DIDSON, and thus precludes any turning near the target for data to be used in identification. Because forward motion causes less image degradation, slow straight-line flyovers of the target that results in a contiguous series of frames can be useful. Hovering maneuvers at different ranges (and hence grazing angles) are also desirable.

Two flyovers perpendicular to each other followed by a hovering maneuver around the object and covering 360 degrees is a recommended start for working with near-range imagery that contains navigation data. Very shallow grazing angles do not need to have complete shadows, but it helps to have some imagery at steeper grazing angels on the same side/view of the object that contains a complete shadow for height estimation.

Frame rates higher than the 3 fps and/or slower forward motion than 2 m/s will help ensure that the target is captured in at least one image in the center of the frame.

Other imagery of interest for identification work would include sequences of frames during a 90-degree roll of the DIDSON while trained on an object. The roll motion should be very slow for optimal results, starting with the DIDSON in the conventional position and ending at the 90-degree position. This would provide data on objects with the fine azimuthal resolution of the sonar interrogating both the horizontal

and vertical dimensions of the observed objects, with many varying stages in between during the roll.

Also important for any data set is supporting documentation about the mission. Water temperature, fresh or salt water, and estimated sound speed are environmental variables that are potentially helpful. Also useful would be information about the target types available and their locations (latitude and longitude), sea conditions, and any other visual imagery (photos on land) and/or dimensions of objects. Also useful would be the period of time the targets have been underwater and the condition of the objects reported by divers through visual/tactile inspection.

When acquiring identification data sets, we have found that the following settings and guidelines for data collection are the most useful:

- high-frequency (HF) setting for greatest resolution
- 4.5-m or 9-m window length (long enough to obtain complete shadows)
- smooth, slow movement, especially in heading rate change, but also in forward speed
- "panoramas" of the area using the LF setting (i.e., when "mowing the lawn") are useful, but not necessary

For future data collections, it would be useful to have the following platform status fields populated in the frame headers of the DDF files:

- latitude
- longitude
- depth
- altitude
- velocity
- heading
- heading rate
- pitch
- roll

Also, the following fields could be useful for data analysis and reconstruction:

- pitch rate and roll rate
- yaw rate
- vector representation of velocity (three spatial dimensions)
- vector representation of acceleration (three spatial dimensions)

REPORT DOCUMENTATION PAGE

Form Approved
OPM No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	November 2005	Technical Report

4. TITLE AND SUBTITLE
A Screening Application for Image Data Collected by an Acoustic Lens Sonar

5. FUNDING NUMBERS
N00014-01-G-0460,
Delivery Order 0005

6. AUTHOR(S)
J.B. Hsieh, J.I. Olsonbaker, and W.L.J. Fox

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Applied Physics Laboratory
University of Washington
1013 NE 40th Street
Seattle, WA 98105-6698

8. PERFORMING ORGANIZATION REPORT NUMBER
APL-UW TR 0502

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Dr. Charles Loeffler
Applied Research Laboratories
The University of Texas at Austin
PO Box 8029
Austin, TX 78713-8029

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

High-resolution, forward-looking sonar systems are critical tools on autonomous underwater vehicle (AUV) platforms where they are used to detect, classify, and localize objects. Data collected by these sensors can be processed automatically and used for navigation (object avoidance) or detailed object examination and discrimination tasks. Once an object is detected, size and shape parameters can be estimated based on the acoustic shadow cast behind the object when ensonified by such sensors. This report discusses the development of adaptive algorithms for highlight localization and shadow segmentation from 1.8-MHz data collected by an acoustic lens imaging sonar—DIDSON (Dual-frequency IDentification SONar). The sonar was mounted in a forward-looking configuration on a REMUS AUV. Algorithms screen large data sets for interesting frames; they demonstrate considerable effectiveness for target detection when the target images are central to the data frame and where there is little platform motion.

14. SUBJECT TERMS
adaptive algorithms, dynamic thresholding, normalization, signal processing, image processing, information processing, high-resolution forward-looking sonar data, autonomous underwater vehicle, AUV, object detection

15. NUMBER OF PAGES
33

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR