2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
Madrid, Spain, October 1-5, 2018

Multibeam Data Processing for Underwater Mapping

Pedro V. Teixeira*, Michael Kaess†, Franz S. Hover*, and John J. Leonard*

Abstract— **From archaeology to the inspection of subsea structures, underwater mapping has become critical to many applications. Because of the balanced trade-off between range and resolution, multibeam sonars are often used as the primary sensor in underwater mapping platforms. These sonars output an image representing the intensity of the received acoustic echos over space, which must be classified into free and occupied regions before range measurements are determined and spatially registered. Most classifiers found in the underwater mapping literature use local thresholding techniques, which are highly sensitive to noise, outliers, and sonar artifacts typically found in these images. In this paper we present an overview of some of the techniques developed in the scope of our work on sonar-based underwater mapping, with the aim of improving map accuracy through better segmentation performance. We also provide experimental results using data collected with a DIDSON imaging sonar that show that these techniques improve both segmentation accuracy and robustness to outliers.**

I. Introduction

From wide-area bathymetric charts used as navigation aids to high-resolution maps of submerged archaeological sites, maps of underwater scenes have become critical to many applications. Depending on the end-use of these maps, there may exist associated accuracy and resolution requirements: in hydrographic surveys, for instance, the features of interest are likely to be at least an order of magnitude larger than those in archaeological or salvage surveys [10]. The accuracy of these maps is driven mainly by the individual accuracies of each of three contributors: *(i)* the pose (position and orientation) estimate, *(ii)* the mapping sensor measurements, and *(iii)* the knowledge of the relative poses between navigation and mapping sensors, also known as *sensor offsets*. It is often the case that the uncertainty associated with the pose estimate is the dominant error source, as evidenced by both the body of literature dedicated to the problem of improving underwater vehicle navigation, and the amount of resources dedicated to improving that estimate in most vehicle platforms through better navigation sensors: high frequency Doppler velocity logs, low-drift inertial measurement units and, in some cases, acoustic positioning systems such as long and ultra-short baseline systems. Still, there are situations in which mapping sensor performance—the main focus of this paper—can be the dominant error source, particularly for platforms with high accuracy navigation systems through use of either highly accurate navigation sensors or more advanced navigation techniques such as Simultaneous Localization and Mapping (SLAM). Additionally, for platforms equipped with less precise navigation sensors, mapping sensor error may exceed navigation error in short-term regimes.

The most common types of sensors used for underwater mapping are optic and acoustic. Optic sensors, such as cameras, structured light sensors, and LIDAR, are capable of very high resolution measurements ($\leq 0.01m$), but their sensitivity to the optical conditions of the water column (turbidity and color attenuation) often places an upper bound on their useful range on the order of a few meters. Lack of ambient light in some applications (particularly those in deep water) requires camera-based mapping sensors to be accompanied by a lighting system.

Multibeam sonars are of particular interest as mapping sensors as they provide a measurement of echo intensity over range and angle, which can be transformed into a set of range measurements. If the pose of the sensor and the nominal orientation of its beams are known, these ranges can be spatially registered and used to create a three dimensional map. Like other sonars, multibeam sensors sit on a trade-off curve between range and resolution, determined by their bandwidth and center frequency: high frequency sonars can achieve very high range resolution, but the higher sound absorption at those frequencies severely limits their range; conversely, lower frequency sensors have long ranges, but lower resolutions. Another important aspect impacting the performance of a multibeam sonar is its beam pattern—it determines the sensor's angular resolution in both horizontal and vertical directions.

The use of multibeam echo sounders for underwater mapping dates back a few decades [11]. Featureless sonar-based mapping techniques rely on classifying sonar images into "object" and "free-space" regions (segmentation) to produce estimates of range to objects, which are then spatially registered using the latest pose estimates. For this reason, range measurement accuracy is highly dependent on segmentation performance. Most approaches rely on some type of thresholding of the sonar image, selecting, for each beam, either the first or strongest return above a threshold. Burguera *et al.* explored different approaches to segment images from a mechanically-swept imaging sonar using standard image processing operations [5]. Similar techniques were used in work by Mallios *et al.* [13], and McVicker *et al.* [14]. Some of these implement some form of post-segmentation outlier rejection, usually by looking at the agreement of the range measurements of adjacent beams [16], [13].

This work was partially supported by the Office of Naval Research (under grants N00014-11-1-0688, N00014-16-1-2628, N00014-16-1-2103 and N00014-16-1-2365) and by the National Science Foundation (under grant IIS-1318392), which we gratefully acknowledge.

*Department of Mechanical Engineering, MIT, 77 Massachusetts Ave, Cambridge, MA 02139, USA {pvt, hover, jleonard}@mit.edu

†Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA kaess@cmu.edu

The use of a fixed threshold in combination with scene-dependent intensities and sonar artifacts results in inconsistent segmentation performance, negatively affecting the mapping process. In his work on underwater mapping [20], VanMiddlesworth highlighted many of the challenges in the use of imaging sonar for mapping, and the difficulties of mitigating their effects on individual images. Instead, he argued for thresholding as an adequate segmentation approach and that mitigation processes (such as outlier removal) should take place later in the mapping pipeline. Other applications of sonar image segmentation appear to take the opposite perspective, and point to more robust processing techniques. Early work by Clay and Stanton [17], [18] looked at using sonar echo statistics for fish and seafloor classification; their work showed that certain differences in the scene (e.g. the density of a fish school) resulted in significantly distinct echo statistics, and proposed the use of the Rayleigh and Rice distributions to model them. In the context of submarine detection, Abraham, Dugelay, and Laterveer [6], [12] addressed the problem of reducing the number of potential detections (*clutter removal*) in a low-frequency sonar ping. Intensity distributions were estimated using the ping in combination with the output from a Page detector [2]. The segmentation problem was then formulated as a Markov random field, which was solved using these distribution estimates and some manually set parameters. Later work by Abraham *et al.* [1] addressed the problem of obtaining these distribution estimates without the use of a detector, relying instead on *expectation-maximization* techniques to determine the parameters of a mixture model.

The aim of this paper is to combine the use of intensity distribution estimation techniques with more robust formulations of the problem of segmentation of multibeam images, while also addressing important pre-processing steps that improve performance of both fixed-threshold and probabilistic classifiers. The paper is structured as follows: we begin with a short overview of the active sonar equation which we use to more formally state our problem in section II. Sections III and IV address techniques to reduce sonar- and channel-induced distortions in the intensity measurements. Section V presents the technique used to estimate a mixture model for the measurements, and how it is used in the segmentation process. These are then demonstrated in section VI using data collected during a ship hull inspection field experiment. We conclude with an overview of the techniques, their strengths and limitations, and avenues for future improvements.

II. Problem Statement

In its simplest form, an active sonar transmits an acoustic signal at some time t_0 and then samples the received acoustic signal over some time interval $[t_0+\delta, t_0+\delta+RT_S]$, where δ is the blanking time, T_S is the sampling period, and R is the number of samples. Assuming no occlusions and a constant sound speed of c, the k^{th} sample of the received signal will have been reflected by an object at a range $r_k \frac{c}{2}(\delta + kT_S)$, or, equivalently, in the k^{th} range bin. The received signal intensity RL for that sample can be modeled using the (active) *sonar equation*:

$$RL = SL - 2TL(r_k) + TS + DI \tag{1}$$

Here, the source level SL is the intensity of the transmitted acoustic signal at the source, while the directivity index DI models the variation in that intensity with direction—also known as its beam pattern. Both parameters are specific to the actual sonar, with data sheets often reporting the beam width as the angle for which the index reaches -3dB (*i.e.*, intensity is halved). The transmission loss $TL(r)$ is a function of medium properties such as the spreading and absorption, as well as the distance between the scatterer and the sonar. Finally, the target strength TS will depend on the the object's dimensions and scattering properties [19]. All of these quantities are expressed in *decibels* with respect to a reference pressure (traditionally, $1\mu Pa$: $dB\ re\ 1\mu Pa$[1]). Since the received signal intensity is quantized, each of the samples of the received acoustic signal will take values in the set $\{0, \ldots, L-1\}$—the output will be a one-dimensional array of quantized intensity values indexed by range.

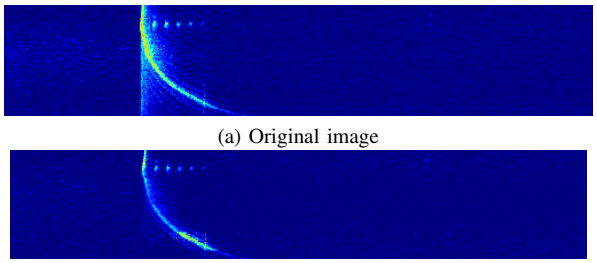
A multibeam sonar, comprising an array of transducers, usually employs beamforming to create a set of virtual beams that sweep along a plane, each of which can be thought of as a single-beam sonar as described above[2]. The output of a multibeam sonar will be a two-dimensional array (image), representing quantized intensity measurements over a polar grid, with each row corresponding to a virtual beam: a sonar with B beams and R range bins (or equivalently, samples over time) will output an $Y_{B\times R}$ array, where y_{ij} represents the quantized average intensity of the received signal for the i^{th} beam's j^{th} range bin.

Our problem can then be described in the following way: we are interested in finding a function $f(\cdot)$ that maps a sonar scan image Y onto a segmented image X. Each pixel y_{ij} in the sonar scan image ($i \in \{1, ..., B\}$, $j \in \{1, ..., R\}$) takes a quantized intensity value in the set $\{0, ..., L-1\}$, whereas each pixel x_{ij} in the segmented image is assigned one of N classes. In the special case when $N = 2$ we have a binary classification problem where we are interested in determining whether or not an object is present at the physical location of pixel x_{ij}. For $N > 2$ we would be looking for a more detailed classification of what is at each pixel (e.g., air bubbles, fish, acoustic artifacts, etc.).

Any classifier will depend on the intensity measurements, so it is critical that these are as accurate as possible. Ensuring this is the case requires identifying—and mitigating—the primary mechanisms through which inaccuracies are introduced. These are primarily related to properties of the sonar and its surrounding medium.

[1]For the sake of simplicity, we will omit the reference pressure ($1\mu Pa$) and distance ($1m$) in the remainder of this text, writing simply "dB".

[2]Some multibeam sonars employ acoustic lens assemblies to accomplish beamforming [3], while some others make use of a 2D array to sweep the beam in two directions and produce a volumetric scan. For simplicity, we focus our discussion on typical imaging and bathymetric multibeam sonars, comprising a one-dimensional linear transducer array



(a) Original image

(b) After deconvolution and taper removal

Fig. 1. Artifacts caused by certain beam pattern properties (side lobes and taper) result in the "ghosting" effect seen in the top image, where multiple instances of the hull shape are noticeable; these can be corrected for using the techniques described in Section III.

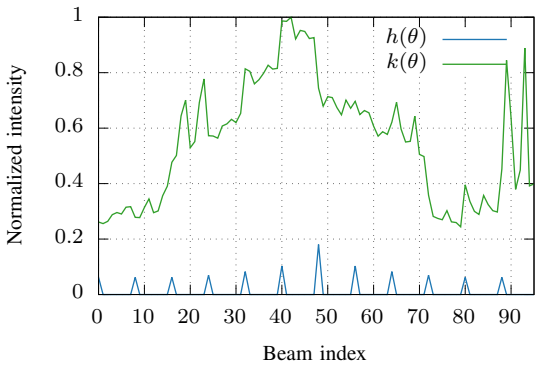


Fig. 2. Angular impulse response and taper function for the DIDSON. The angular impulse response $h(\theta)$ was determined from the sonar specification and normalized to have unit energy; the taper function $k(\theta)$ was determined experimentally from the background set described in section VI.

III. Sonar Model

One of the terms of the sonar equation introduced in the previous section is the directivity index DI, which models the sonar's beam pattern—its directional sensitivity. The beam width of a sonar, defined by the angle at which the main lobe is $3dB$ below its maximum value, determines its angular resolution. The side lobes determine in what other directions is the sonar the most sensitive—depending on the difference in amplitude between main and side lobes, reflections originating from these directions will be more or less noticeable, introducing inaccuracies in the measurements and (further) reducing the overall resolution of the sonar. The effect of these mechanisms is illustrated in Figure 1, where multiple, fainter instances of the hull shape are clearly visible, as a result of significant side lobe height. spatially invariant system $h(r,\theta)$ under the effect of additive noise ν: In order to mitigate these effects, we can model the sonar image $y(r,\theta)$ as the output of a linear, spatially invariant system $h(r,\theta)$ under the effect of additive noise ν:

$$y(r,\theta) = h(r,\theta) * x(r,\theta) + \nu \tag{2}$$

In this model, $x(r,\theta)$ is the original image, also in polar coordinates (as a function of range r and angle θ). The spatial invariance assumption equates to requiring the beam pattern to be the same for all beams and range bins. If we limit the model described by equation 2 to its angular component ($h(r,\theta) = h(\theta)$), the impulse response corresponds to the sonar's beam pattern. Using this model, we can employ Wiener deconvolution to recover an estimate of the original image [9]:

$$\hat{x}(r,\theta) = g(r,\theta) * y(r,\theta) \tag{3}$$

$$G(\omega_\theta) = \frac{H^*(\omega_\theta)}{|H(\omega_\theta)|^2 + SNR(\omega_\theta)^{-1}} \tag{4}$$

$H(\omega_\theta)$ and $SNR(\omega_\theta)$ are, respectively, the power spectral densities of the impulse response and signal to noise ratio.

In this discussion, we have made the assumption that the beam pattern shape—the location, width and magnitude of lobes and nulls—remains the same for all beams. While this is not always the case, we can at least mitigate part of this variation by introducing an angle-varying gain $k(\theta)$ to compensate for the reduction in gain that can be observed in the beam pattern as we move away from the center beam. This is illustrated in figure 2, where there is a clear reduction in response amplitude towards either side of the center beam.

IV. Channel Model

The transmission loss term in Equation 1, TL, captures the attenuation the acoustic signal suffers as it travels from the sonar to the scatterer and back. This reduction is due to two phenomena affecting the signal: *geometric spreading* over an increasingly large surface as it travels in a dispersive medium, and *absorption* by the medium as some of the energy in the pressure waves is converted to heat.

A. Geometric Spreading

Geometric spreading models the decrease in intensity as the signal wave front propagates away from the source along a surface that is increasing in area. Depending on the sonar and environment, this spreading is often modeled as *spherical* or *cylindrical*, with $1/r^2$ and $1/r$ dependencies, respectively [19]. The transmission loss due to geometric spreading, in $dB.m^{-1}$, can be written as:

$$TL_S = 10\log_{10}\left(r^\beta\right) \tag{5}$$

where we have used a reference distance of $r_{ref} = 1$m, and the parameter β models the spreading regime ($\beta = 1$ for cylindrical, $\beta = 2$ for spherical).

B. Absorption

Absorption losses model the decrease in signal intensity caused by dissipation as some of the energy in the signal is converted to heat. This conversion is driven mainly by viscous and relaxation processes related to the water molecules and some of the dissolved substances (respectively). These losses are modeled as:

$$TL_A = \alpha r \tag{6}$$

where α represents the *total* absorption, $\alpha = \alpha_V + \alpha_R$, due to both viscous and relaxation processes. These are mainly dependent on sonar *depth* and *frequency*, as well as water

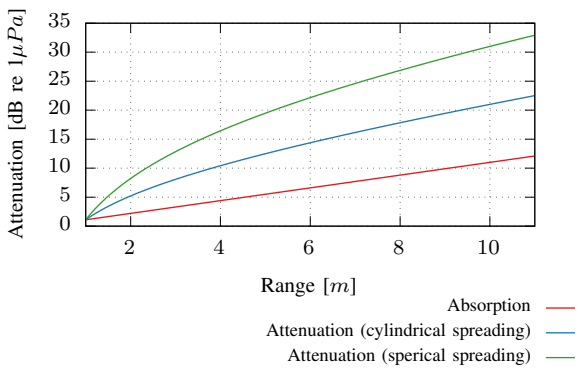


Fig. 3. Attenuation curves for a 1.8MHz sonar under both cylindrical and spherical spreading regimes, using $\alpha = 1.1\ dB.m^{-1}$ [19], [4].

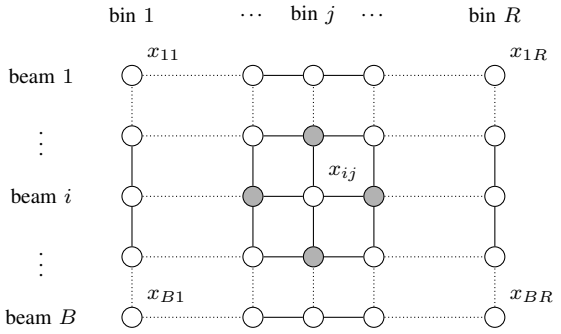


Fig. 4. Graphical model for the classification problem: a $B \times R$ lattice, where each site x_{ij} represents the range-angle cell (pixel) corresponding to beam i and range bin j. The sites making up the 4-neighborhood of x_{ij} are shaded in gray.

salinity and *temperature*, and models for these are available in the literature [19, Ch. 5], [4, Ch. 5]. These attenuation effects are often corrected with a *time-varying gain* (TVG) on the sonar, or as a radial gain on the image [16]:

$$k(r) = k_1 \log_{10}(r) + k_2 r + k_3 \tag{7}$$

where, from equations 5 and 6, we would set $k_1 = 20\beta$ and $k_2 = 2\alpha$. The necessity of ensuring attenuation is compensated for prior to segmentation is illustrated in figure 3, which shows attenuation values under different spreading regimes for the sonar used in section VI.

V. Segmentation

In section II we formulated our problem as one where we want to assign a label to each pixel as a function of the associated intensity measurement y_{ij}. Sections III and IV presented some of the most important phenomena affecting those intensity measurements, as well as techniques that can be used to mitigate them in order to obtain more accurate intensity estimates. The outcome of these techniques is shown in figure 1, where most of the artifacts in the raw sonar image (top) have been successfully removed (bottom), resulting in a more accurate sonar image.

In this section we focus on using these improved estimates for the purposes of segmentation: we begin by estimating the underlying distributions, which are required to characterize the performance of a simple, local binary classifier. We then move to a more advanced model where the label assigned to a site depends not only on the intensity at that site, but also on those of its neighbors.

Our problem formulation described each multibeam image as an image $Y_{B\times R}$ where y_{ij} is the intensity of beam i for the j^{th} range bin. We model this image as a $B \times R$ lattice, as shown in figure 4, where each site (i, j) is treated as a hidden Markov model, with measurement y_{ij} (the intensity of a pixel in the sonar image), and (hidden) state $x_{ij} \in \{1, \ldots, N\}$, describing the scattering at that site.

A. Intensity Distribution Estimates

Treating each site x_{ij} as independent of all other states, we can write the probability distribution associated with its measurement as a N-component mixture:

$$p_{ij}(y) = \sum_{k=1}^{N} \pi_{ij,k} p_{ij}(y|k) \tag{8}$$

where $\pi_{ij,k} \in \{0, 1\}$ and $\sum_{k=1}^{N} \pi_{ij,k} = 1$ for all sites (i, j) (i.e., $\pi_{ij,k}$ is the indicator variable for $x_{ij} = k$). The independence assumption, while clearly incorrect[3], is used to significantly simplify our approach, and is commonly made further along in a mapping pipeline [7]. We have also implicitly assumed that transmission loss, directivity, and all other effects unrelated to the presence (and type) of scattering occurring at site (i, j) have been modeled and corrected.

Under this formulation, the conditional intensity distributions $p_{ij,k}(y|k)$ are free to have different parameters for different sites, radically increasing the dimensionality of the estimation problem, even for single-parameter distributions. To address this, we make the additional assumption that these distributions are identical for all sites:

$$p_{ij}(y) = \sum_{k=1}^{N} \pi_{ij,k} p(y|k) \tag{9}$$

At this point, the conditional distributions are still unknown; to estimate them, we average equation 9 over all sites (i, j):

$$p(y) = \frac{1}{BR} \sum_{i,j} p_{ij}(y) = \sum_{k=1}^{N} \overline{\pi}_k p(y|k) \tag{10}$$

where $\overline{\pi}_k \in [0, 1]$ and $\sum_{k=1}^{N} \overline{\pi}_k = 1$. If $p(y|k)$ are assumed to follow parametric distributions, the estimation problem can be framed and solved using techniques such as expectation-maximization or least-squares, to obtain estimates for both $\overline{\pi}_k$ and the distribution parameters. The choices for $p(y|k)$ are tied to the application, as the scattering properties depend on the environment—options include exponential, Rayleigh, and Rice distributions [18], [1]. Whatever the choice, however, the distribution described by equation 10 should not be treated as stationary (except,

[3]At the very least, the intensity measurement y_{ij} will depend on the presence or absence of an object in sites $(i, 1 \ldots j-1)$.

perhaps, for static scenes): as the scene changes, so will the parameters, and the mixture model must be re-estimated for every new image. To compare $p(y)$ with the empirical distribution $h(y)$, we use the Kullback-Leibler divergence:

$$KLD(p,h) = \sum p(y) log\left(\frac{h(y)}{p(y)}\right) \tag{11}$$

B. Local classifier: binary hypothesis testing

In the special case where $N = 2$, equation 10 becomes

$$p(y) = \overline{\pi}_1 p(y|1) + (1 - \overline{\pi}_1) p(y|2) \tag{12}$$

Under this model, the *maximum a-posteriori* (MAP) decision rule is [15]:

$$\Lambda(y) \underset{x=1}{\overset{x=2}{\gtrless}} \eta \tag{13}$$

where $\Lambda(y) = \frac{p(y|2)}{p(y|1)}$ is the likelihood function, and $\eta = \frac{\overline{\pi}_1}{1-\overline{\pi}_1}$. Here, we set $x = 2$ if $\Lambda(y) > \eta$, and $x = 1$ otherwise. The fixed thresholding decision rule, commonly found in the literature, can be written as:

$$y \underset{x=1}{\overset{x=2}{\gtrless}} \epsilon \tag{14}$$

In this binary hypothesis setting, the probabilities of *false alarm*, p_{FA} and detection, p_D are also of interest:

$$p_{FA} = 1 - F(y|1) \tag{15}$$

$$p_D = 1 - F(y|0) \tag{16}$$

where $F(y|k)$ is the conditional cumulative distribution function. Plotting p_D against p_{FA} yields the *Receiver Operating Characteristic* (ROC) curve, which helps understand classifier performance and the trade-off between the two probabilities. Estimating these is also important when using occupancy grid techniques later in the mapping pipeline [7].

C. Neighborhood classifier: Markov random field

One of the main drawbacks of the local classifier is that it fails to account for information in the neighboring sites, considering only the local intensity value. As illustrated in figure 4, an alternative is to also consider the label of sites x_{kl} in the neighborhood of x_{ij}, v_{ij} when computing the label assignment to x_{ij}. This is done by including the joint label distribution in the model

$$p(x_{ij}, y_{ij}) = \frac{1}{Z} \prod_{ij} \Phi(x_{ij}, y_{ij}) \prod_{kl \in v_{ij}} \Psi(x_{ij}, x_{kl}) \tag{17}$$

and taking its logarithm, yielding

$$E(x_{ij}, y_{ij}) = -ln(Z) + \sum_{ij} \phi(x_{ij}, y_{ij}) + \sum_{kl \in v_{ij}} \psi(x_{ij}, x_{kl}) \tag{18}$$

where Z is the partition function [6], $\phi(x_{ij}, y_{ij})$ is the *unary* potential (the mixture model previously described), and $\psi(x_{ij}, x_{kl})$ is the *binary* potential, often chosen to be a weighted indicator function on the difference. Techniques used to solve for the labels $X = \{x_{ij}\}$ include Iterated Conditional Modes and Simulated Annealing [6], [12].

VI. Experimental Results

A. Experimental Setup

The data used in our experiments was collected using a DIDSON imaging sonar [3], mounted on a Bluefin HAUV [8]. The DIDSON was equipped with a concentrator lens, reducing its vertical beam width from 14° to approximately 1.5°, and operated in *High Frequency* mode, with a center frequency of 1.8 MHz. The data used for the development of these techniques consists of two sets, selected from a larger ship hull inspection set of almost $42,000$ images:

- "Background" set: 116 images of open water (i.e. empty space). Care was taken to ensure these did not contain clutter or targets (e.g. fish).
- "Test" set: a smaller set of just 10 images, representative of possible scenes (i.e. fish, hull, propeller). and shown in the left column of figure 7.

Both subsets were acquired with a receiver gain of 20 dB, and a 9 meter window starting at 2.25 meters.

B. Intensity Distributions

Background: The *background* set was selected from the larger hull inspection set in order to better understand the intensity distribution when no objects are present. The images in this set were pre-processed using the techniques described in section III to recover more accurate intensity estimates. We then computed the intensity histograms for each of the range bins, to determine the need to correct for the phenomena described in section IV—the results are shown in figure 5, where we can see that the variation in intensity due to range is minimal: approximately 5% of full scale at high intensities, and negligible at the more probable, lower intensities. The overall distribution (over all ranges and images) is nearly exponential, save for the higher probability at zero intensity. In fact, when fitting a model of the form

$$p_Y(y|B) = (1 - \pi_B)\delta(y) + \pi_B \lambda e^{-\lambda y} \tag{19}$$

we arrive at $\pi_B \approx 0.312$ and $\lambda \approx 44.6$, with a KL divergence of 4.06×10^{-4}. Here, $\delta(x)$ is the indicator function ($\delta(y) = 1$ when $y = 0$, $\delta(y) = 0$ otherwise). The use of an exponential distribution is consistent with some of the results found in the literature [1].

Object: The images in the *test* set were selected to provide representative examples of the scenes of potential interest in a mapping application, for example: different hull geometries, propeller, fish schools. While each of these objects will certainly have distinct scattering properties, resulting in significantly different intensity distributions, we make the simplifying assumption of treating them as different instances of a particular distribution (namely, the Rayleigh distribution - equation 20), and estimating its parameter for every image. This *scene-dependence* is thus captured in the distribution's prior, and scale parameter σ.

$$p_Y(y|O) = \frac{y}{\sigma^2} e^{-\frac{y^2}{2\sigma^2}} \tag{20}$$

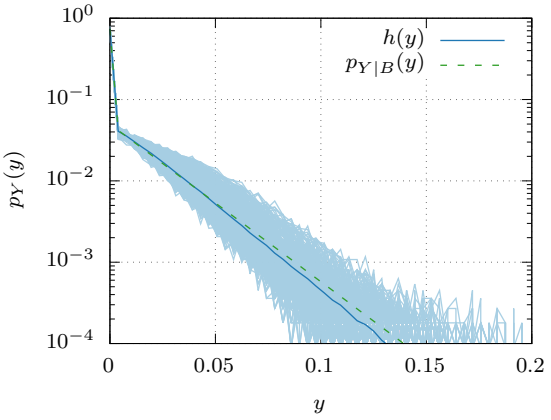


Fig. 5. Background intensity distribution model: each light blue line is the histogram for a single range bin—notice the small variation in intensity ($\sim 5\%$ of full scale); the blue line corresponds to the aggregate histogram over the entire data set. The background intensity distribution (dashed green) is approximated as an exponential distribution with zero bias (equation 19) with $\pi_B \approx 0.312$ and $\lambda \approx 44.6$, for a KL divergence of 4.06×10^{-4}.

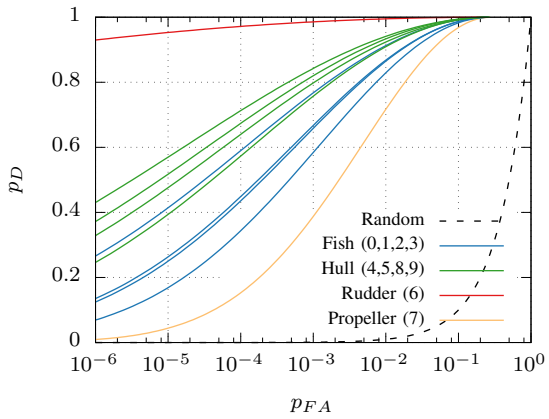


Fig. 6. ROC curves for the test images, with random classifier shown as a dashed black line. Detection probability p_D for $p_{FA} = 0.001$ ranges from approximately 0.4 (propeller) to 0.95 (rudder), with a more typical values of 0.6 (fish) and 0.8 (hull). The two extremes (rudder and propeller) correspond to the worst mixture model fits. For $p_{FA} \approx 1 \times 10^{-5}$ (an average of one false positive per image) p_D drops to 0.5 or lower.

Test Set

Combining equations 19 and 20 we get the mixture model:

$$p_Y(y) = (1 - \pi_1 - \pi_2)\delta(y) + \pi_1 \lambda e^{-\lambda y} + \pi_2 \frac{y}{\sigma^2} e^{-\frac{y^2}{2\sigma^2}} \tag{21}$$

Least-squares fits of this mixture model to each the histograms of the ten test set images (after pre-processing) yields the mixture models shown in figure 8, with parameters listed in table I. From these, we find that the choice of a Rayleigh-distributed object intensity seems appropriate for most cases, with σ usually in the interval $[0.1, 0.2]$ and $\pi_2 \in [0.01, 0.02]$. The background distribution parameters remain, for the most part, consistent with those estimated for the background data set: $\pi_1 \approx 0.3$ and λ in the range 30 – 50. These parameters are also helpful in estimating the signal to noise ratio in the images, which we approximate by $SNR \approx \frac{\pi_2}{1-\pi_2} \approx 0.02$. Still, there are some notable exceptions, namely, images 6 and 7, with the latter achieving the largest KL divergence value over the entire set. In the first case, the faint returns from the partial ensonification of the rudder result in an incorrect fit with $\pi_2 \approx 0$. In the latter—image 7—the object intensity distribution appears to be closer to being exponentially-distributed, likely related to the higher complexity of the scene.

C. Classifier Performance

After estimating the mixture model parameters for the test set, we use equations 15 and 16 to compute the ROC curve for each image. From the results, shown in figure 6, we notice some scene-dependence in the classification performance: for a false positive probability $p_{FA} \approx 1 \times 10^{-5}$ (an average of one false detection per sonar image), the detection probabilities range between 0.2 and 0.4 for the "fish" images (0, 1, 2, and 3), and between 0.4 and 0.6 for "hull" images (4, 5, 8, and 9). It is also interesting to note that images 6 and 7 obtain the best and worst curves for the set, as a result of poor model parameter estimation.

The third column of figure 7 shows the segmentation results using MRF segmentation given an initial label assignment provided by the local MAP classifier (equation 13). We employ Simulated Annealing to estimate the optimal assignment from an energy minimization formulation of equation 18, and use a weighted indicator function to model label mismatch as having a higher energy cost:

$$\psi(x_{ij}, x_{kl}) = \frac{1}{2}(1 - \delta(x_{ij} - x_{kl})) \tag{22}$$

For comparison, the second column of figure 7 shows the output of a fixed threshold classifier, with $\epsilon = 0.22$. The threshold was chosen in order to keep p_{FA} under 1×10^{-5} for typical hull images (an average of 1 false positive per image) - thus requiring knowledge of the underlying distributions. Still, fixed threshold results show significantly worse performance - images 0 and 5 fail to detect a large part of the returns, likely due to the lower intensity measurements from low incidence angles; image 9 shows some outliers despite missing part of the hull.

TABLE I
TEST SET: MIXTURE PARAMETERS

Image	π_1	λ	π_2	σ	KLD $[\times 10^{-3}]$
0	0.292	58.521	0.021	0.131	1.158
1	0.301	42.013	0.015	0.149	0.764
2	0.294	46.006	0.016	0.133	0.565
3	0.294	39.358	0.014	0.138	0.610
4	0.298	50.565	0.011	0.176	1.289
5	0.291	73.630	0.020	0.101	0.665
6	0.312	52.937	0.000	1.000	4.990
7	0.279	26.978	0.070	0.152	7.522
8	0.312	35.356	0.006	0.275	1.612
9	0.310	44.012	0.010	0.191	1.550

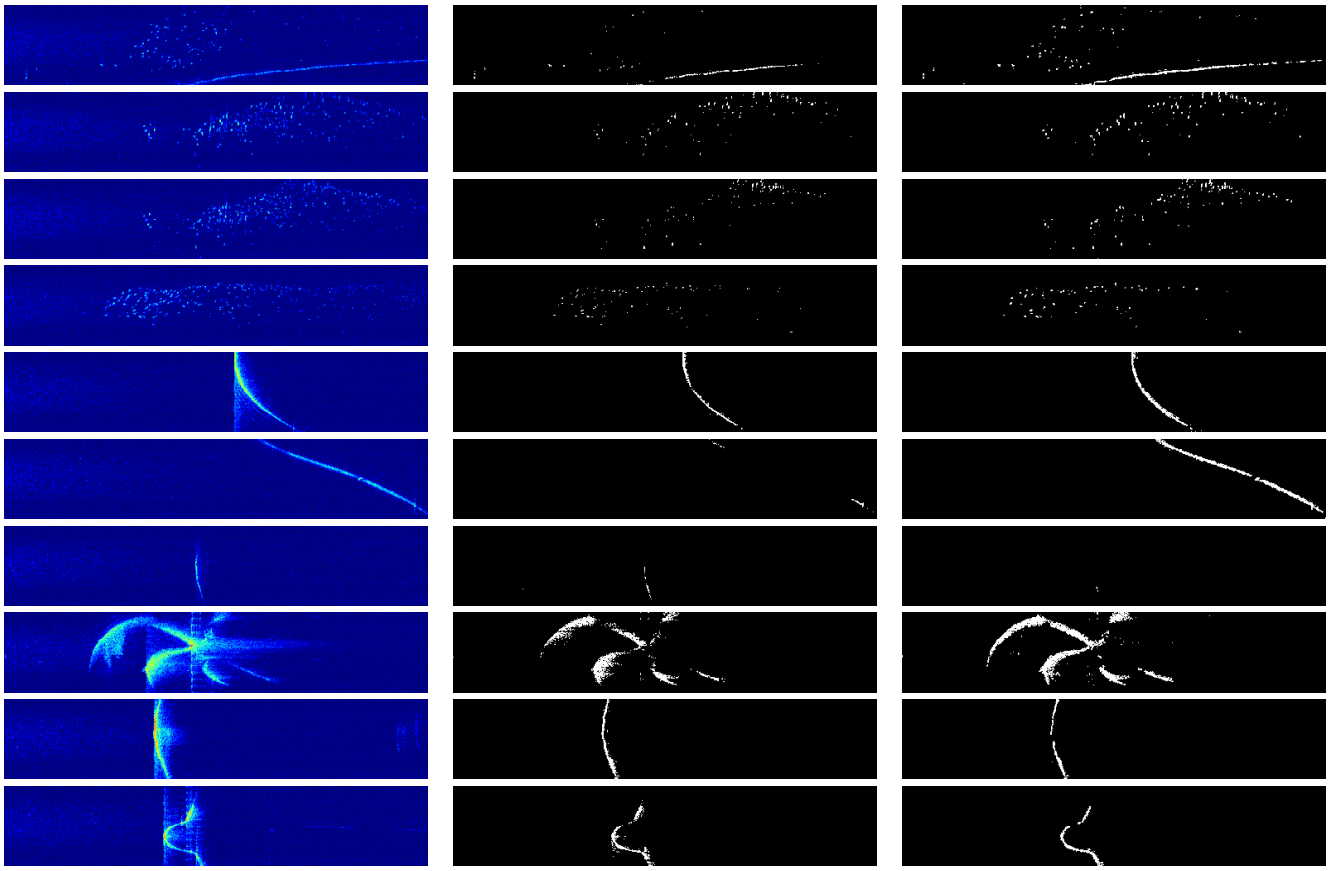


Fig. 7. Test data set results: original test images (first column), fixed threshold (second column), and MRF segmentation using local MAP segmentation for initialization (third column). The test images show several representative scenes for the ship hull inspection task (hull (4,5,8,9), propeller (7), and rudder (6)), as well as other scenes which are often encountered during inspection: fish schools (0,1,2,3) and seafloor (0). The fixed threshold was set to $\epsilon = 0.22$ in order to obtain $p_{FA} \approx 1 \times 10^{-5}$ for the hull images (hence assuming knowledge of the underlying distributions). For the segmented images, background and object are shown as black and white, respectively. All images are shown in polar coordinates, and were pre-processed to mitigate beam pattern and taper effects prior to segmentation. Best viewed on a computer.

VII. Conclusions and Future Work

Sections III and IV provided a brief overview of some of the key phenomena affecting the intensity measurements output by a multibeam sonar. Mitigating sonar artifacts proved particularly effective despite reliance on strong assumptions which may not always hold (e.g. the angular component of the impulse response will not be invariant when imaging occurs in the sonar's near-field). Future efforts could look at relaxing some of these assumptions while working to capture the radial component of the impulse response function. A related limitation of the current work also worthy of future investigation is that it does not account for echo intensity variation with incidence angle (e.g. model scattering as Lambertian)—doing so would likely increase the accuracy of the mixture model and, in turn, that of the classifier, as a result of greater separation between background and object.

The distribution estimates presented in section VI appear to be accurate approximations of the empirical distributions, providing some experimental support for the component choices for the mixture model. Still, images 6 and 7 illustrate some of the limitations: a mismatch likely caused by the inability of the simple two-component mixture model to capture a great variety of scenes, which could be mitigated by either increasing the number of components or considering other distributions (e.g. Weibull). Comparison with other mixture model estimation techniques, such as expectation-maximization, would also be desirable, to better assess feasibility for real-time applications.

While the MRF model has proven successful at capturing some of the dependence between neighboring sites to produce a more accurate segmentation, there is significant room from improvement (as can be seen from the number of outliers still found on the images in the last column of figure 7), potentially through use of larger neighborhoods. Conversely, it would also be interesting to explore simpler model formulations so as to reduce the computational effort, or to learn linear approximations to these accurate, but resource-intensive techniques.

Acknowledgements

The authors would like to thank Nick Rypkema, Mei Yi Cheung, Dehann Fourie, and the anonymous reviewers for their helpful comments and feedback on earlier versions of this manuscript.

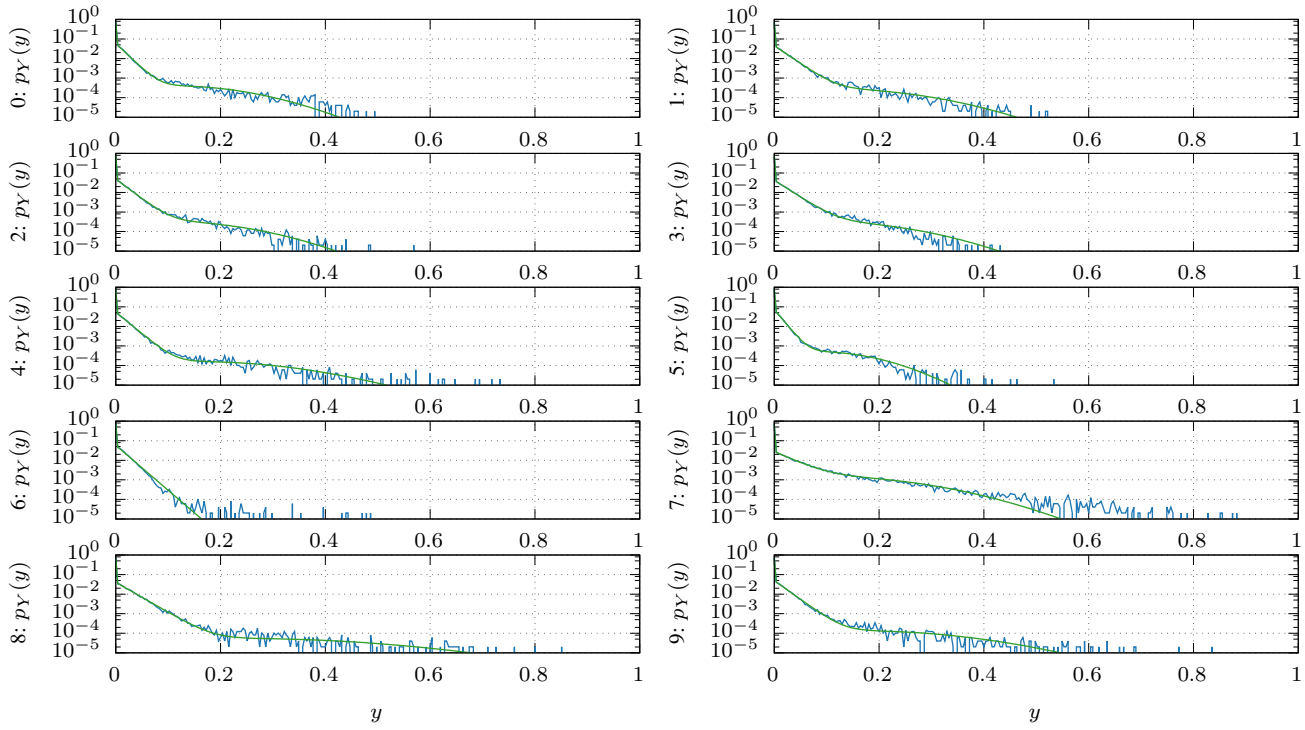


Fig. 8. Mixture models for the test set images: normalized histogram is shown in blue, and mixture model in green. The parameters for these distributions are listed in table I. Plots are numbered left-to-right: top row corresponds to images 0 (left) and 1 (right). Best viewed on a computer.

REFERENCES

[1] D. A. Abraham, J. M. Gelb, and A. W. Oldag, "Background and clutter mixture distributions for active sonar statistics," *IEEE Journal of Oceanic Engineering*, vol. 36, no. 2, pp. 231–247, April 2011. [Online]. Available: http://ieeexplore.ieee.org/document/5773654/

[2] D. A. Abraham and P. K. Willett, "Active sonar detection in shallow water using the page test," *IEEE Journal of Oceanic Engineering*, vol. 27, no. 1, pp. 35–46, Jan 2002. [Online]. Available: http://ieeexplore.ieee.org/document/989883/

[3] E. Belcher, W. Hanot, and J. Burch, "Dual-frequency identification sonar (didson)," in *Proceedings of the 2002 Interntional Symposium on Underwater Technology*. IEEE, 2002, pp. 187–192. [Online]. Available: http://ieeexplore.ieee.org/document/1002424/

[4] W. S. Burdic, *Underwater acoustic system analysis*, A. V. Oppenheim, Ed. Prentice-Hall, 1984.

[5] A. Burguera, G. Oliver, and Y. Gonzlez, "Range extraction from underwater imaging sonar data," in *2010 IEEE 15th Conference on Emerging Technologies Factory Automation (ETFA 2010)*, Sept 2010, pp. 1–4. [Online]. Available: http://ieeexplore.ieee.org/document/5641076/

[6] S. Dugelay and D. A. Abraham, "Reduction of low frequency active sonar clutter through image processing," SACLANT Undersea Research Centre, Tech. Rep., December 1997, sACLANTCEN report SR-272-UU.

[7] A. Elfes, "Using occupancy grids for mobile robot perception and navigation," *Computer*, vol. 22, no. 6, pp. 46–57, June 1989. [Online]. Available: http://ieeexplore.ieee.org/document/30720/

[8] F. Hover, J. Vaganay, M. Elkins, S. Willcox, V. Polidoro, J. Morash, R. Damus, and S. Desset, "A vehicle system for autonomous relative survey of in-water ships," *Marine Technology Society Journal*, vol. 41, no. 2, pp. 44–55, June 2007.

[9] E. E. Hundt and E. A. Trautenberg, "Digital processing of ultrasonic data by deconvolution," *IEEE Transactions on Sonics and Ultrasonics*, vol. 27, no. 5, pp. 249–252, Sept 1980. [Online]. Available: http://ieeexplore.ieee.org/document/1539261/

[10] "Special publication no. 44," International Hydrographic Organization, February 2008, IHO Standards for Hydrographic Surveys.

[11] W. K. S. Jr., "Multisensor modeling underwater with uncertain information," MIT Artificial Intelligence Laboratory, Tech. Rep. 1143, July 1988. [Online]. Available: http://hdl.handle.net/1721.1/6980

[12] R. Laterveer, "Single ping clutter reduction: segmentation using markov random fields," SACLANT Undersea Research Centre, Tech. Rep., March 1999, SACLANTCEN report SR-307. [Online]. Available: http://www.dtic.mil/dtic/tr/fulltext/u2/a371858.pdf

[13] A. Mallios, P. Ridao, E. Hernandez, D. Ribas, F. Maurelli, and Y. Petillot, "Pose-based slam with probabilistic scan matching algorithm using a mechanical scanned imaging sonar," in *OCEANS 2009-EUROPE*, May 2009, pp. 1–6.

[14] W. McVicker, J. Forrester, T. Gambin, J. Lehr, Z. J. Wood, and C. M. Clark, "Mapping and visualizing ancient water storage systems with an rov; an approach based on fusing stationary scans within a particle filter," in *2012 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Dec 2012, pp. 538–544. [Online]. Available: http://ieeexplore.ieee.org/document/6491022/

[15] A. V. Oppenheim and G. C. Verghese, *Signals, Systems and Inference*. Pearson, 2016. [Online]. Available: https://www.pearsonhighered.com/program/Oppenheim-Signals-Systems-and-Inference/PGM106873.html

[16] C. Roman, "Self consistent bathymetric mapping from robotic vehicles in the deep ocean," Ph.D. dissertation, Massachusetts Institute of Technology, 2005. [Online]. Available: http://hdl.handle.net/1721.1/39184

[17] T. K. Stanton, "Sonar estimates of seafloor microroughness," *The Journal of the Acoustical Society of America*, vol. 75, no. 3, pp. 809–818, 1984. [Online]. Available: https://doi.org/10.1121/1.390590

[18] T. K. Stanton and C. Clay, "Sonar echo statistics as a remote-sensing tool: Volume and seafloor," *IEEE Journal of Oceanic Engineering*, vol. 11, no. 1, pp. 79–96, January 1986. [Online]. Available: http://ieeexplore.ieee.org/document/1145139/

[19] R. J. Urick, *Principles of Underwater Sound*, 3rd ed. McGraw-Hill, 1983.

[20] M. A. VanMiddlesworth, "Toward autonomous underwater mapping in partially structured 3d environments," Master's thesis, Massachusetts Institute of Technology, January 2014. [Online]. Available: http://hdl.handle.net/1721.1/87791