Measurement Model Development for the Correlation of Imaging Sonar Acoustic Shadows and Bathymetry for ROV Terrain-Relative Localization

Jose Padial*
jpadial@stanford.edu

Shandor Dektor*
sgd@stanford.edu

Stephen M. Rock*†
rock@stanford.edu

*Aerospace Robotics Laboratory, Stanford University
496 Lomita Mall, Rm 250, Stanford, CA 94305, USA
†Monterey Bay Aquarium Research Institute (MBARI)
7700 Sandholdt Road, Moss Landing, CA 95039, USA

***Abstract*—**

This paper details the development of a probabilistic measurement model for the use of sonar imagery as an automated pilot aid for localization of a remotely-operated vehicle (ROV) with respect to an *a priori* bathymetric terrain map. Specifically, acoustic shadows in sonar imagery are correlated with expected visibility images generated from a bathymetry map, where acoustic shadows are significant drops in the sonar image intensity. An expected visibility image is generated for a given ROV position estimate with respect to a stored terrain map by extracting a bathymetry profile along the sonar scan plane and then evaluating a metric termed "differential height" to quantify visibility probability in the pixel space of the sonar image. This paper provides the theoretical foundation behind the approach presented in [1], and further details the tuning of measurement model parameters using ROV sonar image field data collected in collaboration with the Monterey Bay Aquarium Research Institute (MBARI). The position estimation results for one dataset from [1] is repeated in this paper in order to demonstrate localization performance using the proposed approach.

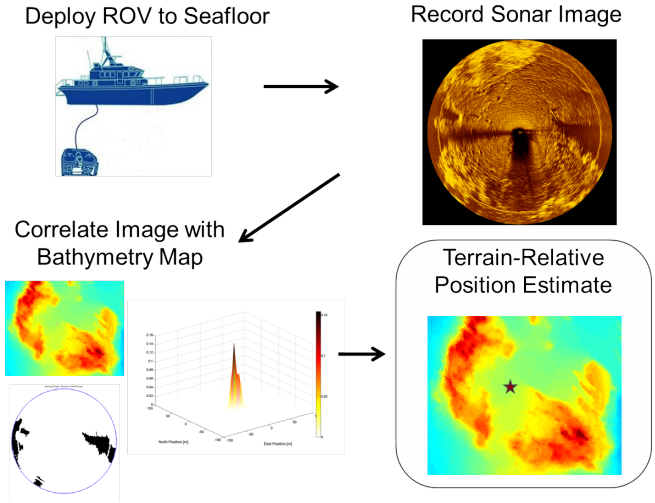


Figure 1: ROV operational motivation for the proposed sonar image-based localization.

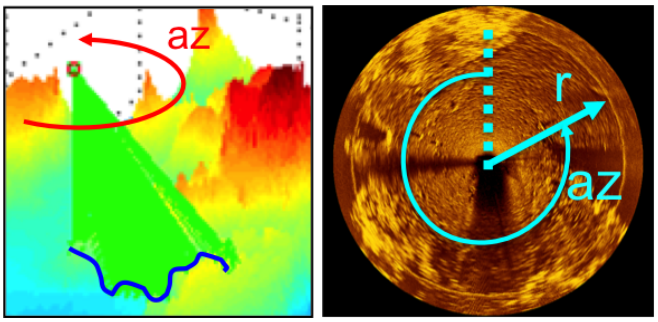
I. Introduction

The ability to localize a remotely-operated vehicle (ROV) accurately with respect to the seafloor terrain enables safer operation and improved navigation to sites of interest. Typical ROV positioning is accomplished through acoustic triangulation of the ROV from the surface vessel, e.g. using an ultra-short baseline (USBL) system. While this method of ROV positioning may be suitable for some mission profiles, it does not provide a terrain-relative position estimate of the ROV, and as such is insufficient for missions requiring accurate knowledge of ROV position relative to the seafloor. A terrain-relative position estimate is particularly imporant when the terrain map is inaccurately geo-referenced, as a USBL latitude/longitude estimate does not account for geo-referencing error. This motivates the development of a terrain-relative position estimator leveraging sensors commonly found on ROVs. In this work, imaging sonars are used.

Imaging sonars measure the intensity versus range for sonar beams over a span of azimuth angles. At each azimuth angle, the sonar transducer emits a pulse of acoustic energy and listens for return pings, much like sidescan sonar systems. Imaging sonars can be broadly classified into multibeam and mechanically-scanned types. Multibeam imaging sonars form separate beams with narrow azimuthal and wide elevation beam-widths using a phased array of transducers. Mechanically-scanned imaging sonars physically rotate the sonar transducer in azimuth. Figure 2 provides a diagram of terrain ensonification and an illustration of the sonar image range-azimuth space.

Sonar imagery is currently used by ROV pilots to infer map-relative position from features in the sonar image. While ROV pilots are skilled at this task, time and effort would be spared if it were automated. Furthermore, an automated process could explore a broader range of possible ROV position hypotheses to account for larger map geo-referencing errors. Figure 1 provides a schematic depiction of the operational motivation for this work.

Using sonar imagery as a quantitative navigation aid requires a measurement model to generate an expected sensor signal as a function of the vehicle pose and map, which can then be correlated with a measured sensor signal to estimate position. Forming a measurement model for sonar image intensity is complex, as intensity returns are functions of grazing angle, terrain surface composition, and water properties [2]. Further, there are gains and filters applied to the raw signal internal to the sensor which are often, as is the case



(a) (b)

Figure 2: Imaging sonar operation. (a) Sonar scan plane (green) for a given azimuth angle ensonifies terrain profile (blue). (b) Sonar image is in a range-azimuth space.

with this work, unknown to the operator. As such, predicting the intensity signal accurately requires more than a vehicle pose estimate and bathymetry map alone. However, *acoustic shadows*, which are significant drops in sonar image intensity, are determined primarily by line-of-sight occlusion due to the geometry of ensonified terrain relative to the sonar transducer, and are therefore well-suited to bathymetry-based correlation techniques.

This paper details the development of a probabilistic measurement model for correlation with acoustic shadows in sonar imagery for ROV localization, and builds off of work presented in [1]. In particular, this paper provides the theoretical foundation for the approach presented in [1], along with details on the tuning of measurement model parameters using ROV sonar image field data. Section II briefly describes related prior work, particularly with respect to the use of sonar imagery for navigation. Section III provides an overview of the localization method. Section IV presents details on the development of a probabilistic measurement model for terrain visibility, and Section V-A details the tuning of measurement model parameters using ROV field data obtained in collaboration with the Monterey Bay Aquarium Research Institute (MBARI). The position estimation results for one dataset from [1] is repeated in this paper in Section V-B in order to demonstrate localization performance using the proposed approach.

II. Related Work

Sonar imagery has been successfully utilized previously for underwater localization, however most methods differ from the approach presented in this paper in that they are based on correlating sonar image to sonar image as opposed to sonar image to bathymetry. These image-to-image methods typically correlate using computer-vision image features, and require a flat seafloor assumption. The flat seafloor assumption is employed due to the inherent ambiguity in extrapolating spatial information from sonar imagery; specifically, the azimuth and range of a given intensity return are known for sonar imagery, but the elevation angle is unknown.

Several examples of image-to-image methods exist. In [3] a method is proposed to detect and match landmarks in sidescan imagery by level set evolution on Haralick feature maps, where the nature of the landmark registration is similar in nature to visual feature matching. The work of Fallon et al. [4] matched sidescan image features across multiple sidescan images and fused the spatial matching with acoustic ranging from a surface vessel for navigation drift mitigation. More recently, the work of King et al [5] explored the extension of computer vision feature technology to sidescan image registration, with results showing successful loop closure registration.

In a similar vein, sonar imagery has been used in AUV Simulataneous Localization and Mapping (SLAM) solutions in the context of landmark detection and matching for improved navigation. In [6], distinctive features are identified in sidescan sonar imagery, and matched across images to estimate and track landmark positions. These landmark correspondences are used to improve the navigation estimate in order to project a more smooth and consistent sidescan map.

The work presented in this paper differs from these past works in that the presented method produces a navigation estimate with respect to an *a priori* terrain map, rather than relative to previous sonar images. Further, this work does not require a flat bottom assumption in order to disambiguate the sonar intensity returns; expected visibility images are generated from a bathymetry map and projected into the sonar image range-azimuth domain, *which is an unambiguous mapping*. In fact, the approach presented in this paper works best for interesting topography, i.e. non-flat terrain.

This approach was previously described by the authors in [7], [1]. In [7] two methods were presented: (1) XOR-based correlation using a deterministic expected shadow signal based strictly on line-of-sight occlusion, and (2) correlation using expected visibility images estimated from the relative heights of terrain with respect to line-of-sight from the sonar transducer, termed "differential height". The differential height method was shown to yield superior navigation performance, and is the measurement model that is used for the localization results presented in this paper. The primary contribution of this paper from [1] is the expanded treatment of the probabilistic measurement model development.

III. Localization Method Overview

The estimation method presented in this paper correlates acoustic shadows in measured sonar imagery with expected visibility images generated using a probabilistic measurement model. Acoustic shadows are large drops in return intensity in the sonar image. While sonar intensity returns from low-reflectance terrain surfaces, e.g. featureless mud, can be low, geometric acoustic shadows have been observed to generate significantly lower return intensities. As such, shadow-based correlation is well suited to bathymetry-based localization.

Acoustic shadows in the measured sonar image are assigned a hard label according to image intensity:

$$M[u,v] = \begin{cases} 0 & \textbf{measured} \text{ shadow} \\ 1 & \text{otherwise} \end{cases} \quad (1)$$

where M is the measured binary image, and (u, v) is the pixel index in the image.

The primary method of labeling shadows is the application of binary thresholding on sonar image intensity. That is, image

pixels below the threshold intensity values are labeled as shadow.

This work is intended for use with non-parametric position estimation filters. Examples of such filters include particle filters and point mass filters (PMFs), both of which are extensively used in underwater navigation solutions, and detailed in [8]. This work focuses on the output of a single measurement update for a PMF. The PMF estimated state is 2-D North and East vehicle position $\mathbf{x} = [x_N, x_E]^T$. The orientation of the vehicle (pitch, roll, heading) is estimated by the ROV inertial navigation system (INS), and altitude is provided by either an altimeter or a doppler velocity logger (DVL).

The output of the PMF is measurement weights for discrete state hypotheses, where a greater measurement weight $w^{(i,j)}$ indicates that position $\mathbf{x}^{(i,j)}$ is more likely. A multi-resolution filtering approach is employed to efficiently estimate ROV position. Initially, a coarse grid of discrete state hypotheses is instantiated about the USBL-estimated ROV position at 4m resolution. Each discrete position hypothesis $\mathbf{x}^{(i,j)}$ in the PMF is assigned a normalized measurement weight $w^{(i,j)}$ according to Equations 2,3.

$$w_0^{(i,j)} = \prod_{(u,v)\in C} \underbrace{M[u,v]P[u,v]^{(i,j)}}_{\text{measured visible}} + \underbrace{(1-M[u,v])\left(1-P[u,v]^{(i,j)}\right)}_{\text{measured shadow}} \tag{2}$$

$$w^{(i,j)} = \frac{w_0^{(i,j)}}{\sum_i \sum_j w_0^{(i,j)}} \tag{3}$$

where M is the measured binary image according to (1), and $P[u,v]$ is the output of the measurement model for the (u,v) pixel of the image. Term C refers to the "correlation region" of the sonar image, which is chosen *a priori* to exclude regions of the image that contain shadows not due to the terrain (e.g. due to ROV geometry).

Higher resolution (1m) grid cells are instantiated about PMF cells with measurement weight above a threshold, and measurement weights are calculated for these higher resolution cells. This multi-resolution PMF thus allows for 1m resolution position estimation in likely areas only, rather than bearing the cost of the 1m resolution estimation over the entire PMF search area.

IV. Measurement Model

The measurement model is a function that outputs an *expected signal* given a vehicle pose estimate and the bathymetric terrain map. An expected signal must be generated for each position hypothesis in the PMF in order to calculate a measurement weight for that hypothesis according to (2),(3). As a first step in the measurement model process, and for a given position estimate, a bathymetry profile of terrain points for each azimuth scan plane of the sonar image is extracted from the stored terrain map. Next, an expected signal for each terrain point in the scan plane is calculated. In order to be correlated with the measured image, the expected signal is then projected into the range-azimuth space of the sonar image. Finally, each pixel (u,v) in the expected signal image is assigned a value through linear interpolation to yield the expected image P from (2).

An expected signal can be deterministic or probabilistic. The first expected signal implemented by the authors was a deterministic binary signal based on line-of-sight occlusion, where the expected signal for terrain point j in the scan plane was given by:

$$P_j = \begin{cases} \lambda & \textbf{expected} \text{ shadow} \\ 1-\lambda & \text{otherwise} \end{cases} \tag{4}$$

where λ must be between 0 and 0.5, and was chosen to be 0.1.

In order to improve position estimation, the authors developed a probabilistic measurement model for terrain visibility with the goal of accounting for uncertainty from map error and unmodeled sonar physics. Section IV-A introduces the visibility probability problem. A multivariate normal distribution (MVN) model for terrain visibility is detailed in Section IV-B, which has a theoretical foundation under Gaussian assumptions. However, the MVN model is shown to be computationally intractable for real-world problems. A simplified "differential height" (DH) model is introduced in Section IV-C to model terrain visibility with limited computational overhead, and in Section IV-D the DH model is shown to be descriptive enough to closely model MVN probabilities for simulated terrain profiles. Parameters of the DH model are then fit with imaging sonar data from ROV field trials in Section V.

A. Visibility Probability Modeling

Terrain visibility probability is the likelihood that a terrain point is visible to an observer. The terrain here is represented as depths at discrete horizontal locations. For simplicity, consider the case of a 1-D lateral dimension (x), and an observer is a sonar transducer located at zero depth, shown in Figure 3.

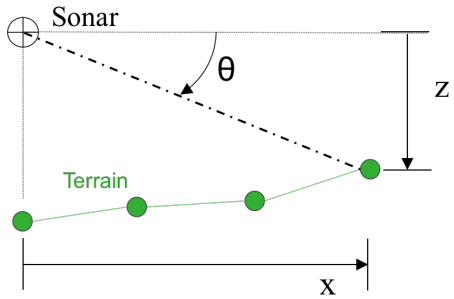


Figure 3: Terrain and sonar observer diagram. Terrain points in green are depths at discrete lateral locations, i.e. a digital elevation map (DEM). A terrain point is specified by lateral distance x and vertical distance z from the sonar transducer, where elevation angle $\theta = \text{arc tan}(z,x)$.

The condition for occlusion of terrain point j caused by terrain point i is given by:

$$\frac{z_i}{x_i} < \frac{z_j}{x_j}, \quad \text{where } i < j \tag{5}$$

where the ratio z_i/x_i is related to the angle from the horizontal to the line-of-sight vector from the sonar transducer to terrain point i by the tangent function.

B. Multivariate Normal Distribution (MVN) Model

In order to properly model terrain visibility, the correlation between terrain points must be modeled. If terrain points are modeled as uncorrelated, for example as univariate Gaussian, then the spatial sampling of the map strongly affects visibility probabilities.

In order to account for correlation between terrain points, terrain depths $\bar{z}$ may be modeled as random variables drawn from a multivariate normal distribution with known mean and covariance. This distribution is conditioned on the map of terrain depths $\hat{z}$, and is given by:

$$\begin{aligned} p(\bar{z}|\hat{z}) &= \frac{p(\hat{z}|\bar{z})p(\bar{z})}{p(\hat{z})} = \eta p(\hat{z}|\bar{z})p(\bar{z}) \\ &\sim \mathcal{N}(\bar{\mu}, \Sigma) \\ &\text{for} \\ \Sigma &= (\Sigma_{\text{map}}^{-1} + \Sigma_{\text{terrain}}^{-1})^{-1} \\ \bar{\mu} &= \Sigma(\Sigma_{\text{map}}^{-1}\hat{z} + \Sigma_{\text{terrain}}^{-1}\bar{\alpha}) \end{aligned} \tag{6}$$

where the terrain prior $p(\bar{z})$ and the map distribution given a terrain $p(\hat{z}|\bar{z})$ are assumed Gaussian as follows:

$$\begin{aligned} p(\bar{z}) &\sim \mathcal{N}(\bar{\alpha}, \Sigma_{\text{terrain}}) \\ p(\hat{z}|\bar{z}) &\sim \mathcal{N}(\bar{z}, \Sigma_{\text{map}}) \end{aligned} \tag{7}$$

The terrain covariance matrix Σ_{terrain} for the prior terrain distribution $p(\bar{z})$ is estimated using tools common to the spatial data analysis community [9]. It is assumed that terrain covariance is stationary, i.e. the covariance between two sample points depends solely on the distance h between the points. This assumption makes the covariance matrix calculable from the covariogram $C(h)$ of the terrain distribution:

$$\begin{aligned} (\Sigma_{\text{terrain}})_{i,j} &= E[(z_i - \alpha)(z_j - \alpha)] \\ &\approx C(h), \quad \text{for } h = ||x_i - x_j|| \end{aligned} \tag{8}$$

where a common mean value α is assumed for all terrain locations. The covariogram of the terrain may be estimated from a terrain map.

The map covariance matrix Σ_{map} can be estimated according to map error estimates. For example, if the creator of a digital elevation map (DEM) specified that the standard deviation of each DEM map cell depth is $0.1m$, then under an assumption of uncorrelation between the map errors between cells map points, the diagonal elements of Σ_{map} could be specified as $(\Sigma_{\text{map}})_{i,i} = (0.1m)^2$.

The probability that point i is visible, denoted p_i, can then be expressed as:

$$\begin{aligned} p_i &= P(\frac{z_1}{x_1} > \frac{z_i}{x_i}, \frac{z_2}{x_2} > \frac{z_i}{x_i}, \ldots, \frac{z_{i-1}}{x_{i-1}} > \frac{z_i}{x_i}) \\ &= P(-z_1 < -\frac{x_1 z_i}{x_i}, \ldots, -z_{i-1} < -\frac{x_{i-1} z_i}{x_i}) \\ &= \int_{z_i^\star = -\infty}^{\infty} \mathrm{F}_{1:i-1}^\star(\frac{x_1 z_i^\star}{x_i^\star}, \frac{x_2 z_i^\star}{x_i}, \ldots, \frac{x_{i-1} z_i^\star}{x_i}) f_i^\star(z_i^\star) \mathrm{d}z_i^\star \end{aligned} \tag{9}$$

where $F^\star(z_1^\star, z_2^\star, \ldots, z_N^\star)$ is the cumulative distribution function (CDF) for the multivariate gaussian random vector $\bar{z}^\star \sim \mathcal{N}(-\bar{\mu}, \Sigma)$, and $F_{1:i-1}^\star(z_1^\star, \ldots, z_{i-1}^\star)$ is the CDF of the marginal gaussian random vector for variables $z_1^\star, z_2^\star, \ldots, z_{i-1}^\star$. The probability density function $f_i^\star(z_i^\star)$ is the marginal distribution for the variable $z_i^\star$.

There are two key problems with practical implementation of this multivariate normal model for estimation of terrain visibility probabilities. The first, and less prohibitive issue, is that the assumptions of the model, particularly with regard to the assumptions of stationarity and a common mean necessary for covariogram use, may not adequately model the true terrain. The more prohibitive problem associated with practical implementation of the multivariate normal model is computational. The use of Equation 9 requires multiple evaluations of a multivariate CDF, which scales in computation time exponentially. This computational burden makes this method of estimating terrain visibility probabilities intractable for any number of points approaching a realistic map. Hence, there is a need for a simplified model that can approximate the terrain visibility probabilities given by Equation 9 with substantially less computational burden.

C. Differential Height (DH) Model

A simplified model is needed for the approximation of terrain visibility probabilities without the prohibitive computational burden of the multivariate normal model.

Toward this goal, a metric termed "differential height" was defined in [1]. The differential height value for terrain point m, δz_m, reflects how occluded or visible the terrain point is. Specifically, the differential height for terrain point m is the distance that m is above or below its occluding line-of-sight, where the occluding line-of-sight is from the transducer to the terrain point j that most occludes m in the scan plane. This is illustrated in Figure 4.

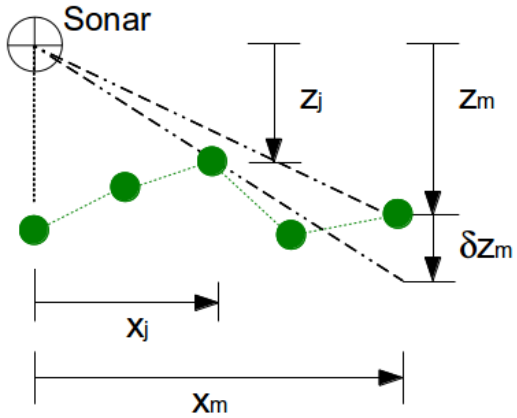


Figure 4: Differential height diagram. Shown is the differential height δz_m of a visible terrain point m above the occluding line-of-sight from terrain point j. Terrain points are shown in green.

If terrain point m is occluded according to line-of-sight from the sonar transducer, δz_m is negative. Similarly, if the

point is visible according to line-of-sight, δz_m is positive. Equation 10 provides the definition of differential height.

$$\delta z_m = \min_j \ (x_m \frac{z_j}{x_j} - z_m), \forall j = 1, 2, \ldots, m-1 \qquad (10)$$

The differential height (DH) model can be viewed as a "one-point" approximation to the MVN model for visibility probability, where the visibility probability in the DH model is estimated according to the most occluding point. This assumes that the true underlying visibility probability for a given terrain point can be adequately estimated solely by a metric derived from its height relation to its most occluding point, rather than with respect to all of the preceding terrain points in the terrain profile (as the MVN assumes).

In order to translate differential heights into visibility probabilities, differential height values are passed through a sigmoid function. The choice of sigmoid function is based in the desire for low visibility probability for low DH values, high visibility probability for high DH values, and a smooth transition region. The sigmoid function chosen is given by the following relation:

$$p_m = \frac{1}{2} + \lambda \frac{\delta z_m - \mu}{\sqrt{\gamma^2 + (\delta z_m - \mu)^2}} \qquad (11)$$

where p_m is visibility probability for terrain point m, μ shifts the mean differential height, and γ and λ adjust the sigmoid shape. Shifting the mean allows for the observed behavior that expected shadows are overly predicted by pure line-of-sight when compared to measured shadows. Setting λ less than the full possible value of ½ accounts for measurement uncertainty, e.g. unmodeled sonar physics, noise and disturbances. Parameter γ adjusts the width of the transition region of the sigmoid. Figure 5 shows the sigmoid given by Equation 11 for $\lambda = 0.4, \mu = -0.2m, \gamma = 0.3m$ (values that are the outputs of the parameter fit to field data detailed in Section V).

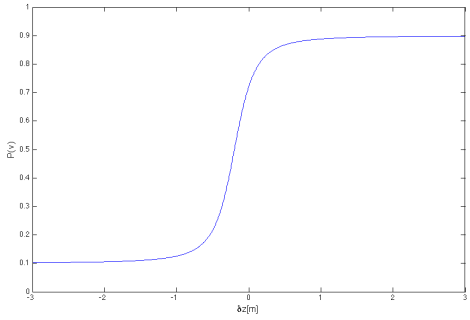


Figure 5: Sigmoid function given by Equation 11 with parameters $\lambda = 0.4, \mu = -0.2m, \gamma = 0.3m$

D. DH Model Fit to the MVN Model

To examine the behavior of the DH model, and determine if it is descriptive enough to adequately approximate the multivariate normal distribution (MVN) model, its estimation of visibility probabilities was compared to that of the MVN model with simulated terrain profiles.

Each simulated profile was a 20-point linear terrain profile $\bar{z}$ sampled from the distribution $\mathcal{N}(\bar{\alpha}, \Sigma_{\text{terrain}})$, where the terrain prior mean $\bar{\alpha}$ was a vector of equal depth values α, and the values in Σ_{terrain} were given by (8) using a covariogram given by the following exponential model commonly employed in spatial data analysis:

$$C(h) = \begin{cases} s & h = 0 \\ (s-a)\exp(-\frac{3h}{r}) & h > 0 \end{cases} \qquad (12)$$

where s is called the "sill" by convention in the spatial analysis literature, a is the "nugget" which specifies a discontinuous drop in covariance at $h = 0$, and r is a specified parameter that describes the level of correlation between terrain points.

The terrain map $\hat{z}$ was drawn from the distribution $\mathcal{N}(\bar{z}, \Sigma_{\text{map}})$. The map covariance matrix was specified as diagonal $\Sigma_{\text{map}} = \sigma^2_{map} I_{20x20}$.

In total, 500 20-point terrain/map profiles were simulated. The parameters α, s, a, r and σ_{map} were varied across profiles in order to diversify the 500 simulated terrains/maps. For each terrain/map profile i, MVN model visibility probabilities P^i_j were calculated for each terrain point j according to (9). Differential heights δz^i_j were calculated for each of the map points. For a given set of DH sigmoid parameters (λ, μ, γ) the DH model visibility probability $\hat{P}^i_j$ for profile i, point j can be calculated by (11).

The parameters (λ, μ, γ) of the DH model were tuned to best fit the MVN probabilities, in a least-squares sense, over the 500 profiles according to the following optimization:

$$\lambda^*, \mu^*, \gamma^* = \underset{\lambda,\mu,\gamma}{\text{argmin}} \sum_{i=1}^{500} \sum_{j=2}^{20} (P^i_j - \hat{P}(\lambda, \mu, \gamma)^i_j)^2 \qquad (13)$$
$$\text{subject to} \quad 0 < \lambda < \frac{1}{2}$$

The optimization yielded values $\lambda^* = 0.5, \mu^* = -0.1m, \gamma^* = 0.2m$. Figure 6 shows the results of the DH model fit for all 500 simulated 20-point terrain/map profiles.

Figure 7 provides the DH model fit to MVN probabilities for the 20-point terrain profile with the median residual error from the optimization given by (13). The middle plot of Figure 7 shows the differential height values for the map points, where negative values reflect points that are occluded according to deterministic line-of-sight in the map. The bottom plot shows the MVN probabilities in red and the DH probabilities in blue. The close alignment of the probabilities shows that the DH model is sufficiently descriptive to closely describe the probabilities given by the MVN model for this terrain profile, while demanding significantly less computation.

V. DH Model Parameter Fit with ROV Field Data

The differential height model fit to the MVN model probabilities shown in Section IV-D is a good theoretical step, but the ultimate use of such a model is for real-world

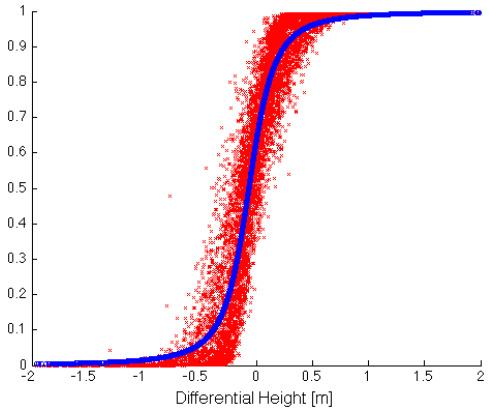


Figure 6: DH model fit to MVN model visibility probabilities for 500 simulated 20-point terrain profiles. MVN probabilities are red, DH model fitted probabilities are blue.

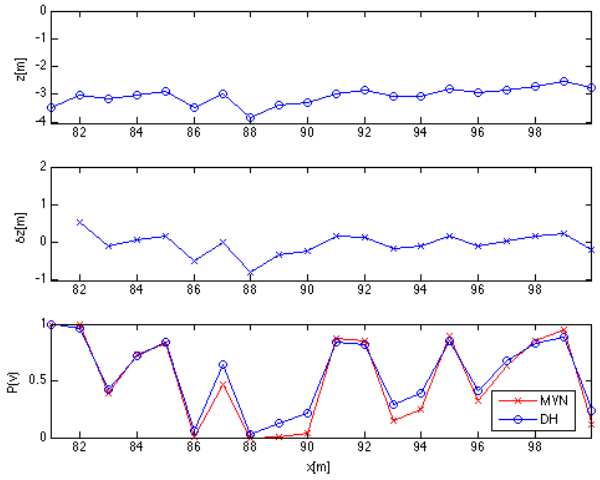


Figure 7: Terrain map, differential heights, and visibility probabilities for the profile with the median residual error from (13). (Top) Map from sonar transducer (sonar at zero depth, zero lateral). (b) Differential heights. (c) Visibility probabilities for the MVN (red) and DH (blue) models.

position localization. As such, the DH visibility probability model should best fit the visibility probability distribution for real data. For the case of underwater localization using imaging sonar, given "true" vehicle locations with respect to underwater terrain maps the DH model can be fit to measured visibility probabilities, which means fitting the parameters of the sigmoid function given by (11) that translates differential heights into visibility probabilities.

Three ROV datasets were collected in the Monterey Bay, in collaboration with the Monterey Bay Aquarium Research Institute (MBARI). The ROVs were outfitted with a Kongsberg mechanically-scanned imaging sonar located on top of the vehicle, as shown in Figure 8. The sonar images obtained have a maximum range of 100m. ROV attitude was measured by a fiber-optic gyro, and altitude was measured by an altimeter. A 1m-resolution DEM was used as the bathymetry map.

Dataset 1 captures the imaging of a large boulder in a crater surrounded by flat terrain at a deep site (roughly 2900m depth). Dataset 2 was collected at shallower depth (86m) in Portuguese Ledge in the Monterey Bay, which is a terrain region with varied topography. Dataset 3 was collected in an area with sand ripples at a deep site near the Dataset 1 location (roughly 2900m depth). For each dataset, a scanning sonar image was collected while the ROV remained motionless.

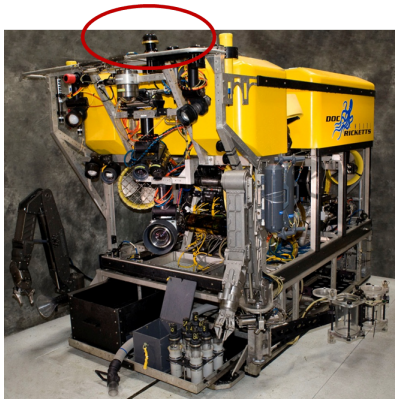


Figure 8: MBARI DocRicketts ROV with a Kongsberg mechanically-scanned imaging sonar circled in red. Image courtesy of mbari.org.

Unfortunately, there was no truth data for these underwater datasets, and subsequently obtaining maximum likelihood estimate (MLE) ROV positions using the Point Mass Filter described in Section III was necessary for the DH model parameter fitting; that is, the MLE positions served as assumed "truth" positions. Using Equations 10, 11 for terrain visibility probabilities, the MLE vehicle seafloor-relative position was obtained for each of the three sonar imagery datasets. Initial values of the sigmoid parameters were iterated until stable MLE positions were found, i.e. the process began with hand-tuned parameters. The differential height values of ensonified terrain about these three MLE positions, along with measured visibility probabilities from the three sonar images, were then used as input to an optimization for sigmoid model parameter estimation. Hence, the parameter identification and MLE estimation were not uncorrelated. There is some circularity in the parameter identification.

A. *Parameter Fit via Nonlinear Least-Squares Optimization*

The left images in Figure 9 are the measured sonar imagery, thresheld on intensity to yield white pixels (visible) and black pixels (shadow). The right images are differential height images for the corresponding MLE locations, where tending toward red indicates higher visibility differential height (more positive), and blue indicates lower visibility differential height (more negative). The large, dark blue regions in the lower parts of the MLE images indicate non-correlation regions, where the differential height values are masked-out in order to prevent consideration of "false" shadows not caused by terrain. The alignment of the measured shadow regions in the left plots of Figure 9 with lower DH values (yellow to green and blue) on the right plots, especially evident in the top two datasets, indicates that these MLE positions are indeed near the true vehicle locations.

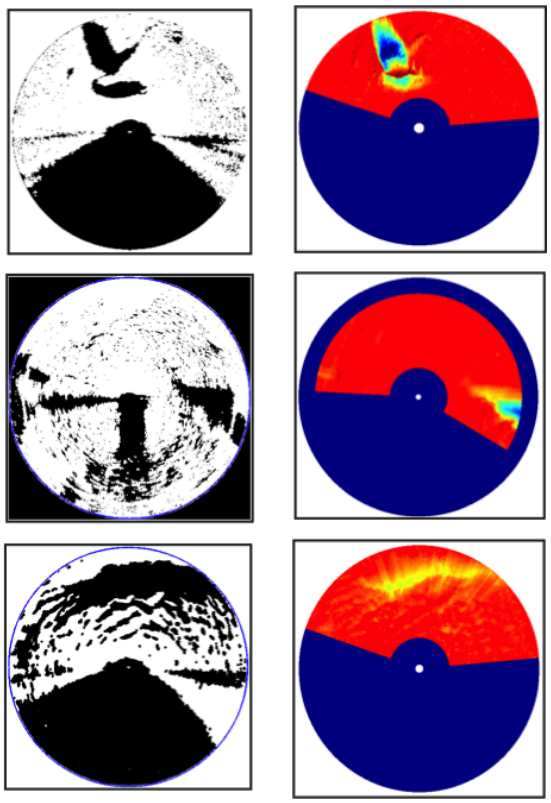


Figure 9: Measured sonar shadow imagery and MLE differential height images. (Left) Measured sonar imagery, thresheld on intensity to yield white pixels (visible) and black pixels (shadow). (Right) differential height images for the corresponding MLE locations, where tending toward red indicates higher visibility differential height (more positive), and blue indicates lower visibility differential height (more negative).

In order to carry out the sigmoid parameter optimization, measured and modeled visibility probabilities were calculated as a function of differential height. First, differential height values were divided into N total bins. For each differential height value in the three MLE differential height images its corresponding DH bin was found, and then the corresponding pixel value in the measured image found. If the pixel was visible (white), then the bin count total and the measured count total for the DH bin were updated, otherwise for a shadow pixel (black) only the bin count total was updated, as depicted in Figure 10.

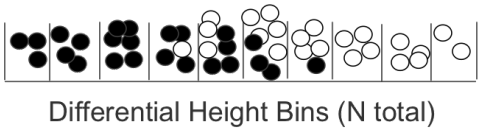


Figure 10: Illustration of measured visibility probability calculation by differential height binning. Shadow pixels (black) and visible pixel (white) for a given measured sonar shadow image are added to the differential height bin for the corresponding MLE differential height image.

The measured visibility probability for the DH value for bin i, P^i, and the modeled visibility probability for DH bin i as a function of the sigmoid parameters (λ, μ, γ), $\hat{P}^i$, are then calculated by:

$$\begin{aligned} P^i &\equiv \text{measured visibility probability for DH bin i} \\ &= \frac{\sum \text{Measured visible pixels for DH bin i}}{\sum \text{Measured pixels for DH bin i}} \\ \hat{P}(\lambda,\mu,\gamma)^i &\equiv \text{modeled visibility probability for DH bin i} \\ &= \frac{1}{2} + \lambda \frac{\delta z^i - \mu}{\sqrt{\gamma^2 + (\delta z^i - \mu)^2}} \end{aligned} \tag{14}$$

The DH sigmoid parameter estimation is then formulated as a weighted least-squares optimization:

$$\begin{aligned} \lambda^*, \mu^*, \gamma^* &= \underset{\lambda,\mu,\gamma}{\text{argmin}} \sum_{i=1}^{N} \beta^i (P^i - \hat{P}(\lambda,\mu,\gamma)^i)^2 \\ &\text{subject to} \quad 0 < \lambda < \frac{1}{2} \end{aligned} \tag{15}$$

where the weight β^i for each DH bin i is set to be its normalized pixel count, capped at some maximum value Q_{max}:

$$\begin{aligned} \beta_0^i &= \min(\sum \text{Pixels for DH bin i}, Q_{\max}) \\ \beta^i &= \frac{\beta_0^i}{\sum_i \beta_0^i} \end{aligned} \tag{16}$$

The DH model sigmoid parameter values obtained through the optimization of Equation 15 are $(\lambda^*, \mu^*, \gamma^*) = (0.4, -0.2m, 0.3m)$. Figure 11 presents the results of the optimization model fit, along with the associated DH bin weights.

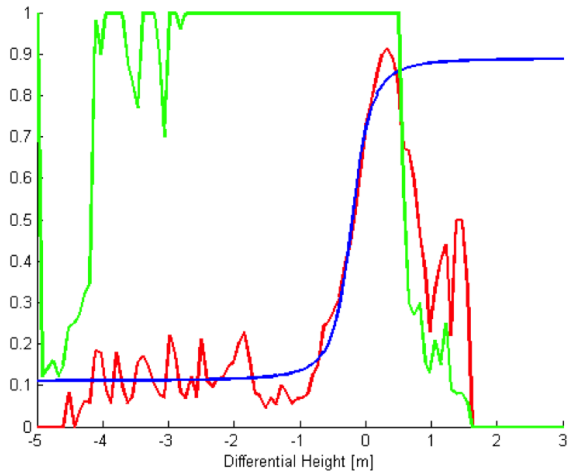


Figure 11: Results of the DH model sigmoid paramter fit. The horizontal axis is differential height [m]. (Red) Measured visibility probabilities. (Blue) Modeled visibility probabilities. (Green) Optimization weights β, scaled on the plot such that the maximum weight is one.

The red line in Figure 11 indicates the measured visbility probabilites, while the blue line shows the modeled probability for the starred (optimized) parameters. The green line shows the weights β, where the weights are scaled for plotting purposes such that the maximum weight is equal to one. Note that the modeled probability fit does extremely well through the transition region from low confidence to high confidence, and acts as essentially a low pass filter on the low confidence measured probabilities (for negative DH values). At higher DH values, the measured probabilities dive lower than expected, though this effect is de-weighted because there were relatively few pixels observed for these DH values. The cause of this drop in measured probabilities for higher DH values is likely a result of noise in shadow labeling.

B. Localization Results

Using the DH model for terrain visibility probability described in Section IV-C with the parameters fit in Section V, ROV position estimation using the Point Mass Filter estimator described in Section III was conducted for the three ROV sonar imagery datasets. These data were previously reported in [1], with localization performance summarized in Table I as was reported in [1]. The new contribution of this paper is the detail on the measurement model development used to generate these localization data. The PMF correlation results for one dataset (Dataset 2) from [1] are repeated below in order to demonstrate the localization performance of the proposed approach.

Set	$\mathbf{x}_{\text{mle}}$	r_{mle}	$\mathbf{x}_{mean}$	r_{mean}	σ_1, σ_2
1	4, 35	35.2	4.0, 34.8	35.0	0.5, 0.4
2	2, 39	39.0	1.4, 39.4	39.4	1.5, 0.7
3	11, 26	28.2	10.7, 27.4	29.4	2.6, 0.5

Table I: Dataset estimator statistics. All values are in meters. For each set, vector quantities $\mathbf{x}_{mle}, \mathbf{x}_{mean}$ and scalar ranges r_{mle}, r_{mean} are measured from the USBL-estimated ROV position for that set. Vector quantities $\mathbf{x}_{mle}, \mathbf{x}_{mean}$ are in local East, North coordinates. The 1σ covariance ellipse semi-major and semi-minor lengths are given by σ_1, σ_2.

Figure 12 presents the PMF correlation results for the Portuguese Ledge dataset (Dataset 2). The values given are normalized measurement weights as calculated by Equations 2,3. The left two plots show the measurement weights for the coarse-scale (4m) resolution of the broad search PMF as described in Section III, where the right two plots show the weights for the fine-scale (1m) resolution. Note the strong unimodal peaks in both the coarse- and fine-grid results, with the correlation peak well-behaved about the MLE estimate.

Figure 13 presents the measured sonar image, measured binary shadow image, MLE visibility probability image calculated using the DH model, and alignment image for Dataset 2. In each binary shadow image, black indicates shadow. In each MLE visibility probability image, blue indicate lower confidence in visibility (i.e. stronger confidence in shadow), with red indicating higher confidence in visibility. The alignment plot was generated by thresholding the expected visibility image at 0.5 to form a binary expected image; red in the alignment plot indicates visibility agreement between measured and MLE images, green indicates shadow agreement, and blue indicates disagreement. The grey hatched region in the MLE visibility probability image and the alignment image is the non-correlation region, i.e. the complement to the correlation region C discussed in Section III.

There is strong agreement between the measured and MLE shadow regions, specifically between the large shadow region on the right side of the image, and the smaller rock-created shadow region on the left, as shown in Figure 13. In the absence of truth data for the ROV position, this strong alignment between the measured imagery and the MLE imagery provides a strong justification for the claim that the MLE position is indeed near the true vehicle location.

VI. Conclusion

This paper detailed the development of a probabilistic measurement model for the use of sonar imagery as an automated pilot aid for localization of a remotely-operated vehicle (ROV) with respect to an *a priori* bathymetric terrain map. The proposed approach could serve as a navigation augmentation to current ROV ship-based USBL acoustic positioning, and could specifically provide a terrain-relative estimate of ROV position that is inherently not afforded by USBL. The theoretical foundation of the proposed approach was newly presented in this paper, along with the tuning of measurement model parameters using ROV sonar image field data.

Acknowledgment

The authors thank the Monterey Bay Aquarium Research Institute (MBARI) for all data and the opportunity to test results at sea. Jose Padial is supported by NASA National Space Technology Research Fellowship NNX11AM92H. Shandor Dektor has received support from a National Defense Science and Engineering Graduate Fellowship and a Stanford Graduate Fellowship. This work was supported in part under NASA ASTEP Grant NNX11AR62G.

References

[1] J. Padial, S. Dektor, and S. Rock, "Correlation of imaging sonar acoustic shadows and bathymetry for rov terrain-relative localization," in *MTS/IEEE OCEANS Conference*, (Taipei, Taiwan), 04/2014 2014.

[2] J. Bell, *A Model for the Simulation of Sidescan Sonar*. PhD thesis, Heriot-Watt University, 1995.

[3] S. Stalder, H. Bleuler, and T. Ura, "Terrain-based navigation for underwater vehicles using side scan sonar images," in *Oceans 2008*, vol. 1-4, pp. 51–53, 2008.

[4] M. F. Fallon, M. Kaess, H. Johannsson, and J. J. Leonard, "Efficient auv navigation fusing acoustic ranging and side-scan sonar," in *ICRA*, pp. 2398–2405, 2011.

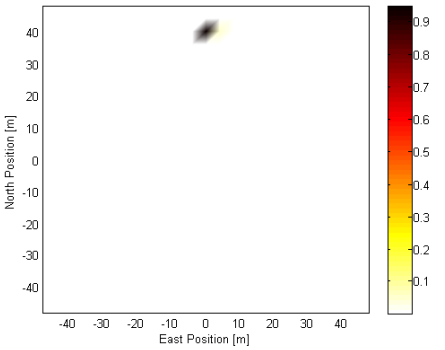
[5] P. King, B. Anstey, and A. Vardy, "Comparison of feature detection techniques for auv navigation along a trained route.," in *MTS/IEEE OCEANS 2013*, pp. 1–8, 2013.

[6] I. Ruiz, Y. Petillot, and D. Lane, "Improved AUV navigation using side-scan sonar," in *OCEANS 2003 MTS/IEEE: CELEBRATING THE PAST...TEAMING TOWARD THE FUTURE*, pp. 1261–1268, 2003.

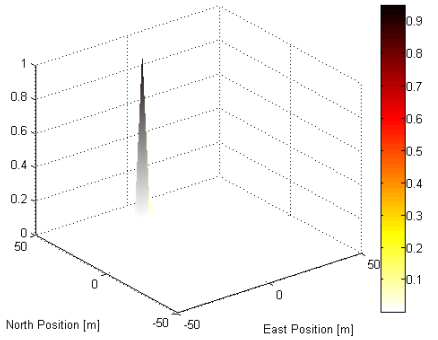
[7] J. Padial, S. G. Dektor, and S. M. Rock, "Correlation of sidescan sonar acoustic shadows and bathymetry for terrain-relative navigation," in *Unmanned Untethered Submersible Technology 2013*, (Portsmouth, NH), September 2013.

[8] K. Anonsen and O. Hallingstad, "Terrain aided underwater navigation using point mass and particle filters," in *Position, Location, And Navigation Symposium, 2006 IEEE/ION*, pp. 1027–1035, 2006.

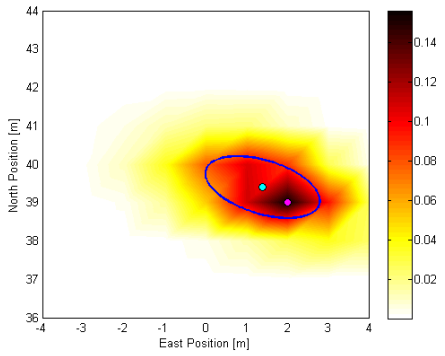
[9] N. A. Cressie, *Statistics for spatial data*, vol. 900. J. Wiley, 1993.



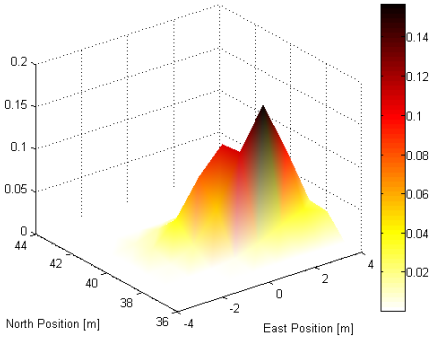
(a) Coarse-grid Top View



(b) Coarse-grid Perspective View

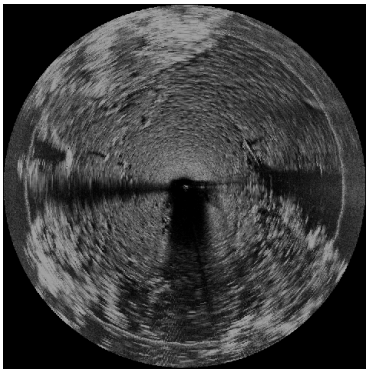


(c) Fine-grid Top View

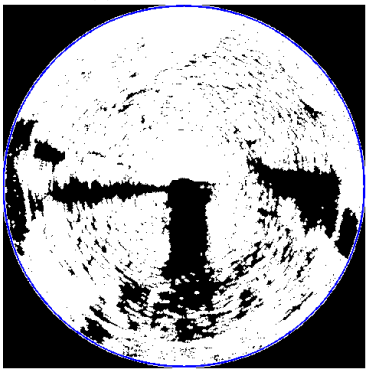


(d) Fine-grid Perspective View

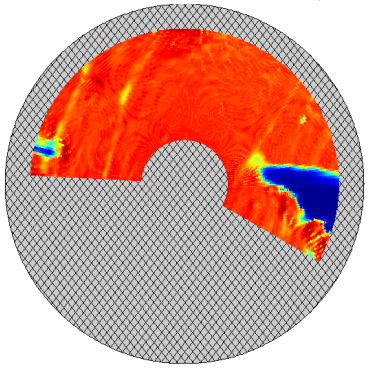
Figure 12: Coarse- and fine-grid PMF correlation results for Dataset 2 (Portuguese Ledge). Correlation results are normalized measurement weights for each position hypothesis. In (c) the magenta point is the MLE position, the cyan point is the mean position, and the blue ellipse is the 1σ confidence ellipse.



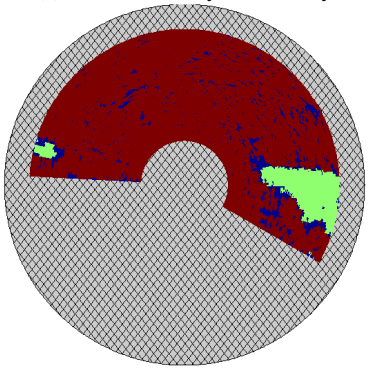
(a) Measured Sonar



(b) Measured Shadow Binary



(c) MLE Visibility Probability



(d) MLE Alignment

Figure 13: Dataset 2 measured sonar image, measured shadow image, PMF MLE expected visibility probability image, and alignment image. Note that in the MLE visibility image, blue indicates higher confidence in shadow, red indicates higher confidence in visibility. In the alignment plot, red indicates agreement between measured and expected visibility, green indicates agreement between measured and expected shadows, and blue indicates disagreement.