2020 IEEE International Conference on Robotics and Automation (ICRA)
31 May - 31 August, 2020. Paris, France

A Volumetric Albedo Framework for 3D Imaging Sonar Reconstruction

Eric Westman, Ioannis Gkioulekas, and Michael Kaess

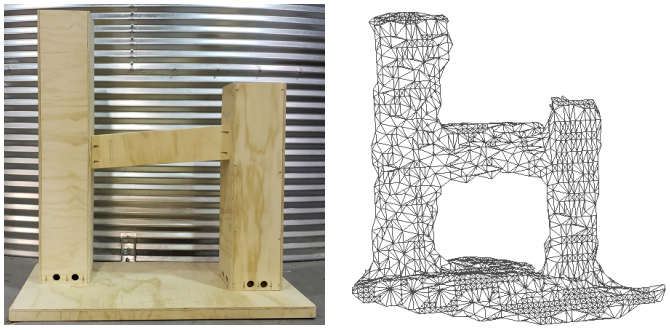
***Abstract*— We present a novel framework for object-level 3D underwater reconstruction using imaging sonar sensors. We demonstrate that imaging sonar reconstruction is analogous to the problem of confocal non-line-of-sight (NLOS) reconstruction. Drawing upon this connection, we formulate the problem as one of solving for volumetric albedo, where the scene of interest is modeled as a directionless albedo field. After discretization, reconstruction reduces to a convex linear optimization problem, which we can augment with a variety of priors and regularization terms. We show how to solve the resulting regularized problems using the alternating direction method of multipliers (ADMM) algorithm. We demonstrate the effectiveness of the proposed approach in simulation and on real-world datasets collected in a controlled, test tank environment with several different sonar elevation apertures.**

I. Introduction

Imaging sonars, also known as forward-looking sonars (FLS), have been widely adopted as a sensing modality for autonomous underwater vehicles (AUVs) and remotely operated vehicles (ROVs) in recent years. These acoustic sensors offer long-range imaging capabilities in turbid waters that are often encountered in the field, where standard optical cameras typically have a limited sensing range. Thanks to the robustness and versatility they provide, imaging sonars have been utilized for a variety of underwater tasks, including localization [18, 23, 28, 37, 38, 39], mapping [4, 5, 14, 16, 17, 24, 25, 32], object detection [19, 20], and planning [15, 29].

Imaging sonars are quite analogous to optical cameras in that they provide 2D image measurements of a 3D environment. Each pixel in an optical image denotes the measured light intensity corresponding to discretized azimuth and elevation angle bins, but the range of an imaged surface is not measured. In contrast, each pixel in an FLS image corresponds to discrete azimuth and range bins, with the elevation angle remaining ambiguous. Some sensors utilize lenses to change the elevation aperture, such as the SoundMetrics DIDSON[1] sensor, which can achieve elevation apertures of $1°$, $14°$, or $28°$.

While techniques such as multi-view stereo (MVS) [27] have enabled dense 3D reconstruction of objects and scenes using optical cameras, this is a difficult goal to achieve with imaging sonars. MVS relies on the principle of photometric consistency – the observation that surface patches tend to look similar from different viewpoints. In the acoustic case, where the sensor actively "illuminates" the scene with sound waves, the appearance of surfaces in sonar images is highly dependent on the viewing geometry. The combination of this image formation model, high noise levels, low bandwidth, and acoustic artifacts combine to make 3D reconstruction from sonar images a challenging problem that cannot be solved by directly adopting standard MVS techniques from computer vision.



(a) (b)

Fig. 1: (a) Photograph of the test structure that was custom made for our test tank experiments. (b) A 3D mesh reconstruction generated from a pointcloud extracted from our volumetric albedo scene representation.

Rather than looking to classical MVS techniques to generate 3D reconstructions with imaging sonar, we draw inspiration from advances in non-line-of-sight (NLOS) reconstruction. This problem consists of reconstructing a scene from indirect active illumination, by looking through a diffusing medium or imaging based on reflections off of a matte surface, such as a wall. The specific contributions of this work are as follows:

- noting the equivalence between imaging sonar reconstruction and confocal NLOS reconstruction for the first time in the literature
- a novel volumetric albedo framework for imaging sonar reconstruction, inspired by recent works in the field of NLOS scene reconstruction
- evaluation of our proposed reconstruction algorithm on simulated and real-world datasets.

The remainder of this paper is organized as follows. We discuss previous approaches to the problem of imaging sonar reconstruction with known poses and their limitations and shortcomings in Section II. In Section III we present the volumetric albedo framework and subsequently in Section IV, the proposed ADMM algorithm to solve for the 3D reconstruction. We demonstrate the effectiveness of our approach on simulated and real-world datasets in Section V and offer concluding remarks in Section VI.

This work was partially supported by the Office of Naval Research under grant N00014-16-1-2103 and by the DARPA REVEAL program under contract HR0011-16-C-0025.

The authors are with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA. {ewestman, igkioule, kaess}@andrew.cmu.edu

[1]www.soundmetrics.com/Products/DIDSON-Sonars/DIDSON-300m

II. Related Work

Efforts to generate 3D reconstructions using sonars actually predate the development of recent FLS imaging sonars. One such early work segmented images from a sector scanning sonar into shadow and object regions. The height of the object of interest is estimated based on the shadow length and a 2.5D elevation map is generated [40]. A similar approach was adopted to build elevation maps from larger-scale side-scan sonar data [22]. These methods are tailored for use in seafloor mapping scenarios, and only succeed in generating elevation maps, rather than full 3D reconstructions.

More recent methods have used imaging sonars to generate rich 3D reconstructions of underwater scenes. In [30], a concentrator lens is fixed to a DIDSON sonar to narrow the elevation aperture to 1°. Assuming that all detected points lie at 0° elevation, a global map of a ship hull scene is reconstructed by registering local submaps using iterative closest point (ICP). While the results are promising for large-scale scenes, the assumption that all imaged points lie in a plane introduces significant error to the reconstruction, and prohibits extending this method to operation with a wider elevation aperture.

Several volumetric algorithms have been proposed for smaller-scale 3D reconstruction. Aykin et al. propose a space carving method in which the entire scene is assumed to consist of occupied space until it is observed to be unoccupied [3, 5]. While this method ideally generates a mesh that bounds the object of interest, many non-convex geometries cannot be accurately reconstructed. The same principle is applied using minimum filtering over a voxel grid to achieve similar results [10, 11]. Another volumetric algorithm [33] builds on the classic occupancy grid mapping framework [9] by distributing occupancy probability to every voxel along each elevation arc. As a probabilistic framework, it is more robust to inaccuracies in the sensor pose than space carving, especially when extended in [34] to include a graph optimization to align local submaps into a unified global map, much like Teixeira et al. [30]. However, occupancy grid mapping considers each voxel independent and does not reason about the relationship between voxels observed by same pixel of a sonar image.

Another line of work has used a generative image formation model based on diffuse reflection to reconstruct the interior of an object given 3D edge initialization using shadow cues [2, 4]. This approach has been extended for use on an AUV outside of a laboratory controlled seafloor mapping scenario [36]. Similarly, Negahdaripour et al. formulate an optimization that refines an initial reconstruction from space carving by attempting to match the image predicted by the generative model with the actual sonar images [25]. In practice, sonar images do not adhere very well to the proposed generative models due to factors such as high noise levels, imprecise sonar calibration, multipath returns, and interference between transducers, making it difficult to generate accurate 3D reconstructions using such methods.

Guerneve et al. [11] approximate the elevation aperture as linear, effectively modeling each pixel's elevation arc as a vertical line segment parallel to the sensor's z-axis. By restricting motion to pure translation along the z-axis, 3D volumetric reconstruction is framed as a blind deconvolution with a spatially varying kernel that captures the surfaces' reflection properties. Assuming uniform reflectivity, the problem is reduced to a linear least squares problem, solved with ℓ_1 regularization by means of an interior point method [21]. This approach is similar to our proposed method, but the linear approximation and sensor motion restriction severely limit its practical application. Our proposed method may be viewed as a generalization of, and improvement over, this approach, as we do not introduce linearization errors or place any restrictions on the sensor motion.

III. Volumetric Albedo Framework

In this section we discuss how imaging sonar reconstruction with known poses is an inverse problem with the same structure as NLOS reconstruction and present a volumetric albedo framework for 3D reconstruction. First, we summarize the important characteristics of NLOS reconstruction.

A. NLOS Reconstruction as Volumetric Albedo

Consider the NLOS scenario of reconstructing a scene around a corner by imaging diffuse reflections off of a wall. A point $\mathbf{l}$ on the line-of-sight (LOS) planar wall is illuminated, often by a laser pulse, at time $t = 0$. Light is scattered in all directions according to some unknown bidirectional reflection distribution function (BRDF), and bounces off of some NLOS surfaces at various points in time, and possibly multiple surfaces at the same instant in time. The third and final bounce occurs when the light reflects off the LOS wall again, and back toward the sensor. The sensor detects light reflected from a distinct sensing point along the LOS wall $\mathbf{s}$, resulting in the so-called 5D light transient $i(t; \mathbf{l}, \mathbf{s})$, which is the intensity (or photon count) as a function of time and points $\mathbf{l}$ and $\mathbf{s}$. Fourth and higher-order bounces are usually ignored due to the difficulty of detecting them and for the sake of simplifying the imaging model.

In the confocal case, the illumination and sensing point on the wall are the same – that is, $\mathbf{l} = \mathbf{s}$, resulting in a 3D light transient as a function of t and $\mathbf{s}$ [26]. Since the speed of light and relative position of the sensor to $\mathbf{l}$ are known, a transient measured at point $\mathbf{l}$ is equivalent to a series of *range-only* measurements taken by a virtual sensor located at $\mathbf{l}$, where the intensities of the measurements are determined by the reflectance properties and geometry of the NLOS surface. The azimuth and elevation angles of the measured surfaces are lost due to scattering. Thus, NLOS reconstruction is an ill-posed inverse problem wherein multiple 1D range measurements must be sampled from a variety of locations on the LOS scene in order generate constraints on the azimuth and elevation angles of the NLOS surfaces.

Due to the complexity of this inverse problem, the scene is often modeled as a volume wherein each point $\mathbf{p}$ is

described by a directionally uniform albedo $\rho(\mathbf{p})$. This simplified model greatly reduces the complexity of the inverse problem, as compared to attempting to account for surface normals and BRDF [13]. Under this assumption, the forward measurement model for the transient in the confocal case is

$$i(t;\mathbf{l}) = \iiint_{\Omega_\mathbf{p}} \rho(\mathbf{p}) \frac{\delta(\mathbf{p} \in \mathcal{S}_{ct})}{\|\mathbf{p}-\mathbf{l}\|_2^4} \, \mathrm{d}\mathbf{p} \tag{1}$$

where $\Omega_\mathbf{p}$ denotes the 3D volume of the NLOS scene, $\mathcal{S}_{ct}$ denotes the sphere of radius ct centered at $\mathbf{l}$, and the quartic term accounts for the spatial propagation of scattered light. For more details on the problem of NLOS reconstruction, we refer the reader to [1, 7, 12, 31].

B. Imaging Sonar Reconstruction as Volumetric Albedo

In the imaging sonar case, each column in an image is analogous to the 1D series of range-only measurements in the NLOS scenario. However, the azimuth angle is disambiguated by an array of transducers, constraining measured surfaces to lie upon a 1D elevation arc rather than a 2D sphere. Thus, under the volumetric albedo model, the forward measurement model for the imaging sonar is given by

$$I(\theta, r) = \iiint_{\Omega_\mathbf{p}} \rho(\mathbf{p}) \, \delta(\mathbf{p} \in A_{\theta,r}) \, \mathrm{d}\mathbf{p} \tag{2}$$

where $A_{\theta,r}$ denotes the 1D elevation arc with limited aperture that corresponds to an image pixel (θ, r). Since the measurement is linear in the albedo, this model may be discretized as a linear system

$$\mathbf{b} = \mathbf{A}\mathbf{x} \tag{3}$$

where $\mathbf{b} \in \mathbb{R}_+^N$ is the vector representation of all N discrete image measurements $I(\theta, r)$ from all images, and $\mathbf{x} \in \mathbb{R}_+^{n_x n_y n_z}$ is the vector representation of the discretized albedo volume, with n_x, n_y, and n_z representing the size of the voxel grid in the corresponding dimensions. $\mathbf{A}$ is a sparse binary matrix that corresponds pixel measurements to voxels that lie on the corresponding elevation arc. If the sonar images are not pre-processed to compensate for the spatial spreading of sound, then a range-based gain may be applied to (2), which would scale entries in $\mathbf{A}$ accordingly. This linear system directly follows from the discretized albedo volumes used for NLOS reconstruction [1, 12, 13, 31]. However, the sonar linear systems are significantly sparser than those in the NLOS scenario, since each measurement corresponds to a 1D elevation arc manifold, rather than a 2D ellipsoidal or spherical manifold.

To compute the correspondence matrix $\mathbf{A}$, a forward projection (projecting center points of voxels into image pixels) or back projection (projecting pixels along their elevation arc into the voxel grid) procedure may be used. For our experiments, we use forward projection, noting that both projection procedures may be parallelized to improve computational efficiency.

Some of the first attempts to solve this large, sparse system in the NLOS case approximated the solution using backprojection:

$$\mathbf{x}_{\mathrm{bp}} = \mathbf{A}^T\mathbf{b}. \tag{4}$$

A commonly used heuristic is to apply a filter after backprojection, such as the Laplacian filter, to sharpen the result [31]. Our proposed method attempts to solve (3) via regularized optimization, which is inspired by similar works in the NLOS literature [12]. Several convolutional approximations have been proposed for the NLOS problem that greatly increase the computational efficiency of these optimization-based approaches [1, 26]. While similar approximations may be made in the imaging sonar case, we find that the sparsity of $\mathbf{A}$ and relatively limited resolution of the sonar sensor make standard optimization procedures much more efficient for imaging sonar reconstruction than for NLOS reconstruction.

A key shortcoming of this framework is that it does not capture the effects of occlusion. One possible way to address this is to only use low-intensity pixels in each image column from the shortest range until the first high-intensity pixel. These correspond to free space under the image formation model. Low-intensity pixels at ranges beyond the first imaged surface may correspond to actual surfaces in the scene that are occluded by surfaces closer to the sensor.

One of the benefits of this framework is that the albedo of the entire scene may be solved for jointly. This contrasts with prior works utilizing occupancy grid mapping [33] or minimum filtering [10], which update each voxel independently. Furthermore, formulating the forward sensing model as a sparse linear system enables the use of convex optimization methods to guarantee convergence to a global minimum.

IV. ADMM Optimization

The linear system derived from the volumetric albedo formulation of imaging sonar reconstruction may be solved using a least squares optimization:

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \, \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2 + \Gamma(\mathbf{x}) \tag{5}$$

where $\Gamma(\mathbf{x})$ is a term that combines all priors or regularization terms. If no regularization is used, the solution is trivial to obtain but may be rather inaccurate. We propose utilizing three separate priors that are commonly used in the NLOS volumetric albedo literature [12] to generate smooth, continuous surface reconstructions: non-negativity, weighted ℓ_1 regularization, and total variation regularization. These may be expressed as:

$$\begin{aligned}\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \, & \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2 + \mathcal{I}_{\mathbb{R}_+}(\mathbf{x}) + \\ & \lambda_1\|\mathbf{W}\mathbf{x}\|_1 + \lambda_{\mathrm{TV}}\|\nabla\mathbf{x}\|_1\end{aligned} \tag{6}$$

This may be reformulated as a separable objective function with linear constraints:

$$\begin{aligned}\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \; & g_1(\mathbf{z}_1) + g_2(\mathbf{z}_2) + g_3(\mathbf{z}_3) + g_4(\mathbf{z}_4) \\ \text{subject to} \; & \underbrace{\begin{bmatrix}\mathbf{A}\\ \mathbf{I}\\ \mathbf{W}\\ \nabla\end{bmatrix}}_{\mathbf{C}}\mathbf{x} - \underbrace{\begin{bmatrix}\mathbf{z}_1\\ \mathbf{z}_2\\ \mathbf{z}_3\\ \mathbf{z}_4\end{bmatrix}}_{\mathbf{z}} = \mathbf{0}.\end{aligned} \tag{7}$$

This may be solved using the alternating direction method of multipliers (ADMM) algorithm, which is often used to solve the volumetric albedo problem in the NLOS literature [12, 13]. Then, the augmented Lagrangian is

$$\mathcal{L}_\rho(\mathbf{x},\mathbf{z},\mathbf{y}) = \sum_{i=1}^{4} g_i(\mathbf{z}_i) + \mathbf{y}^T(\mathbf{C}\mathbf{x}-\mathbf{z}) + \frac{\rho}{2}\|\mathbf{C}\mathbf{x}-\mathbf{z}\|_2^2. \tag{8}$$

We proceed using the notation of scaled ADMM, where $\mathbf{u} = \mathbf{y}/\rho$. We compute the update for $\mathbf{x}$ using gradient descent, which is much faster than inverting $\mathbf{C}$:

$$\mathbf{x} \leftarrow \mathbf{x} - \frac{\rho}{\mu}\mathbf{C}^T(\mathbf{C}\mathbf{x}-\mathbf{z}+\mathbf{u}) \tag{9}$$

where μ controls the step size. The update for each component of $\mathbf{z}$ utilizes the proximal operators corresponding to each $g_i(\mathbf{z}_i)$:

$$\begin{aligned}
\mathbf{z}_1 \leftarrow & \underset{\mathbf{z}_1}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{z}_1-\mathbf{b}\|_2^2 + \frac{\rho}{2}\|\mathbf{v}-\mathbf{z}_1\|_2^2,\ \mathbf{v}=\mathbf{A}\mathbf{x}+\mathbf{u}_1 \\
= & \frac{\mathbf{b}+\rho\mathbf{v}}{1+\rho} \\
\mathbf{z}_2 \leftarrow & \underset{\mathbf{z}_2}{\operatorname{argmin}} \mathcal{I}_{\mathbb{R}_+}(\mathbf{z}_2) + \frac{\rho}{2}\|\mathbf{v}-\mathbf{z}_2\|_2^2,\ \mathbf{v}=\mathbf{x}+\mathbf{u}_2 \\
= & \max(0,\mathbf{v}) \\
\mathbf{z}_3 \leftarrow & \underset{\mathbf{z}_3}{\operatorname{argmin}} \lambda_1\|\mathbf{z}_3\|_1 + \frac{\rho}{2}\|\mathbf{v}-\mathbf{z}_3\|_2^2,\ \mathbf{v}=\mathbf{W}\mathbf{x}+\mathbf{u}_3 \\
= & S_{\lambda_1/\rho}(\mathbf{v}) \\
\mathbf{z}_4 \leftarrow & \underset{\mathbf{z}_4}{\operatorname{argmin}} \lambda_{\mathrm{TV}}\|\nabla\mathbf{z}_4\|_1 + \frac{\rho}{2}\|\mathbf{v}-\mathbf{z}_4\|_2^2,\ \mathbf{v}=\nabla\mathbf{x}+\mathbf{u}_4 \\
= & S_{\lambda_1/\rho}(\mathbf{v})
\end{aligned} \tag{10}$$

where $S_\kappa(a) = (a-\kappa)_+ - (-a-\kappa)_+$ is the soft threshold function. Finally, the dual variable update is

$$\mathbf{u} \leftarrow \mathbf{u} + \mathbf{C}\mathbf{x} - \mathbf{z}. \tag{11}$$

This iterative ADMM procedure is performed until convergence of the primal and dual residuals, as defined in [6].

The entire ADMM optimization is performed during each iteration of an iteratively reweighted ℓ_1 minimization procedure (IRL), in order to further enhance sparsity [8, 12]. We initialize the ℓ_1 weighting matrix $\mathbf{W}$ as identity for the first IRL iteration, and update it at each IRL iteration as

$$\mathbf{W}^{j+1} := \operatorname{diag}\left(\frac{1}{|\mathbf{x}^j|+\epsilon}\right). \tag{12}$$

V. Evaluation

In evaluating our proposed imaging sonar reconstruction framework, we are primarily concerned with the results on real-world datasets. However, we find it helpful to evaluate the results on simulated datasets as well, due to the availability of ground-truth models and sensor poses.

We evaluate our proposed method against backprojection for comparison (denoted ADMM and BP, respectively). BP is an approximate solution to the inverse problem computed using (4). This is a commonly used benchmark solution in the NLOS literature. It is also akin to the occupancy grid mapping (OGM) approach to imaging sonar reconstruction [33, 34]. Both OGM and BP disregard the correlation between voxels corresponding to the same measurement — for each measurement, both methods update each corresponding voxel independently based on some function of the pixel intensity (e.g. inverse sensor model).

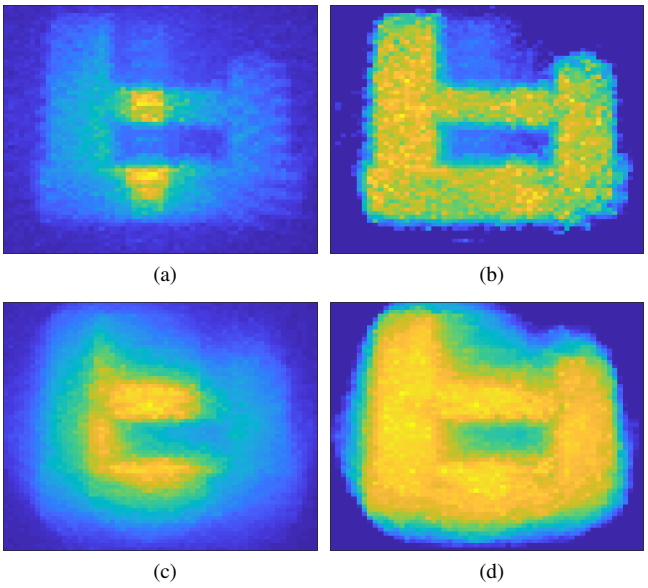


Fig. 2: Maximum intensity projection (MIP) images for (a) BP and (b) ADMM on a simulated dataset with 3° elevation aperture. (c) and (d) show the MIP images for BP and ADMM, respectively, on a simulated dataset with 10° elevation aperture. Blue regions correspond to low albedo, and yellow regions to high albedo.

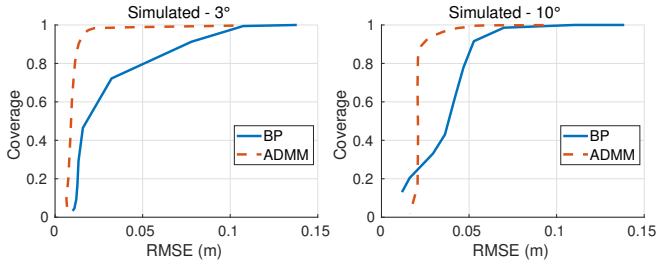


Fig. 3: Coverage vs. error curves for simulated datasets using (a) 3° elevation aperture and (b) 10° elevation aperture.

In all experiments we use a voxel grid with 2.5 cm resolution. For the ADMM optimization, we use $\rho = 1$, $\mu = \|\mathbf{C}\|_2^2$ (estimated by the power method), and $\epsilon = 0.01$.

A. Metrics

For qualitative evaluation, we show the maximum intensity projection (MIP) images of evaluated volumes. A MIP image shows the maximum intensity of all voxels along one particular direction and is a useful tool for visualization.

There are a variety of ways to quantitatively evaluate the accuracy of a volumetric reconstruction against a ground-truth pointcloud. We choose to extract pointclouds from the volume by thresholding the albedo and taking the centers of voxels that exceed the threshold as surface points. The pointcloud is aligned to the ground-truth model using a known transformation for simulated datasets and a manually

tuned transformation for test tank datasets, since ground-truth alignment is not available. We evaluate two metrics using the aligned pointclouds: coverage and error. The coverage is defined as the ratio of points in the ground-truth model for which the closest point in the reconstructed model is within a certain Euclidean distance. We use the length of a voxel diagonal as the threshold when computing coverage. For the error metric, we compute the root-mean-square error (RMSE) of the Euclidean distance from all points from the reconstructed model to the closest ground-truth point. Varying the threshold throughout the feasible range allows for trading off between the coverage and error of the reconstructed pointcloud and yields a curve much like the receiver operating characteristic (ROC) curve of a binary classifier. We present these curves for quantitative evaluation of our reconstructed volumes.

B. Simulation

We simulate sonar images of the structure shown in Fig. 1a, utilizing a ground-truth point-cloud scan scan collected using a Faro Focus 3D laser scanner. We use a set of sonar poses that based on a realistic set of viewpoints that could be imaged using an AUV or ROV. The datasets consist of 180 images with motion between sensor poses limited to the sensor's $x-y$ plane. An image is generated at each of 18 different roll angles induced at 10 different points in the $x-y$ plane, which could be acquired on an AUV by simply translating and yawing the vehicle. Images are generated by projecting all points lying in the sonar field of view into the sonar image, with the intensity of each pixel proportional to the number of imaged points. A more accurate pixel intensity model may also be used. However, the precise intensity as a function of the sensor and surface geometry has minimal effect on our results, as the volumetric albedo framework does not consider surface reflection properties. Surfaces that would actually be occluded are still visible in the simulated images. Finally, normally distributed noise is added to each pixel to simulate the low SNR of real acoustic sensors.

Fig. 2 shows the MIP images for the resulting BP and ADMM volumes for simulated datasets with 3° and 10° elevation apertures. Compared to the BP volumes, the ADMM volumes show much more distinct surfaces and lower intensities in sections that correspond to free space. Naturally, the reconstructions are less precise with greater elevation ambiguity, but the general shape of the structure is still clearly visible in the ADMM MIP for the 10° dataset. Likewise, the coverage vs. error curve for our ADMM reconstruction achieves significantly lower error for the same amount of coverage as BP, as shown in Fig. 3. Note that using a discretized volume representation limits the minimum possible achievable error. In these simulated experiments, we use $\lambda_1 = \frac{\phi_{\text{fov}}}{5}$ and $\lambda_{\text{TV}} = \frac{\phi_{\text{fov}}}{20}$, where $\phi_{\text{fov}} = \phi_{\max} - \phi_{\min}$ is the elevation field of view in degrees. The regularization coefficients ought to increase with the elevation aperture to account for the increased number of voxels observed per measurement.

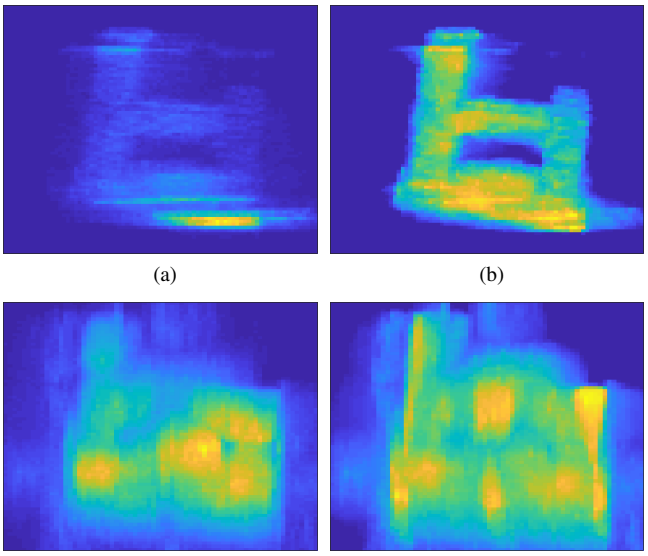


Fig. 4: Maximum intensity projection (MIP) images for (a) BP and (b) ADMM on a test tank dataset with 1° elevation aperture. (c) and (d) show the MIP images for BP and ADMM, respectively, on a test tank dataset with 14° elevation aperture. Blue regions correspond to low albedo, and yellow regions to high albedo.

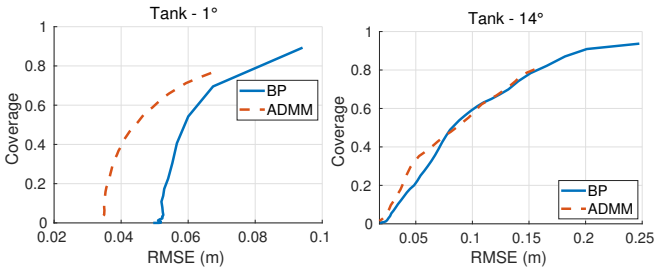


Fig. 5: Coverage vs. error curves for simulated datasets using (a) 1° elevation aperture and (b) 14° elevation aperture.

C. Test tank

We also evaluate our proposed method on real-world datasets collected in a test tank environment using a Sound-Metrics DIDSON imaging sonar mounted on a Bluefin Hovering Autonomous Underwater Vehicle[2] (HAUV). Due to the limited size of the test tank, we keep the DIDSON fixed and pointed directly downward from the vehicle, scanning the submerged structure in Fig. 1a from above. We collect datasets with a concentrator lens to narrow the elevation aperture to approximately 1° and with no lens for a 14° elevation aperture. Over 900 images are required with 1° elevation aperture to achieve full coverage of the structure, and over 400 images for 14° elevation aperture.

The HAUV's onboard odometry measurements are used to provide the pose estimates for both BP and ADMM. Although the odometry measurements are quite accurate due to the combination of a high-end IMU and a Doppler velocity log (DVL), the pose estimate inevitably drifts after prolonged use. To maintain the integrity of the pose estimates for the reconstruction procedures, we limit the length of the datasets

[2]www.gdmissionsystems.com/products/underwater-vehicles/bluefin-hauv

to only a few minutes long, which is expected to limit the pose drift to less than one degree rotation and several centimeters in translation. The vehicle is remotely controlled to perform three to four sweeps parallel to the sensor's z-axis, offset along the sensor's y-axis, to ensure full coverage of the structure.

Fig. 4 shows the MIP images for the $1°$ and $14°$ test tank datasets. Much like the simulated results, the structure is more well-defined and clearly visible in the ADMM MIP images than BP. The coverage-error curves for both resulting volumes are shown in Fig. 5. As in simulation, the ADMM results show marked improvement over BP for the narrow aperture dataset. The difference in performance is not significant for the wide aperture dataset, which is due to the limited set of viewpoints, particularly the lack of roll rotation (around the sonar's x-axis).

Sample pointclouds extracted from both test tank datasets are shown in Fig. 6. The narrow aperture reconstruction is highly accurate and covers the entire top surface of the structure, with only a few outlier points due to multipath reflections from the interior of the hollow structure. The wide aperture reconstruction cannot fully disambiguate the structure due to the lack of rich viewpoints, but still captures the main components: the base, two vertical pilings, and crossbar.

VI. Conclusion

In this work, we have connected the problems of NLOS and imaging sonar reconstruction for the first time in the literature. We have presented an algorithm for imaging sonar reconstruction with known poses that generalizes previous work to arbitrary sensor motion and does not make any linearization approximations. The proposed framework is solved via ADMM and may incorporate a variety of priors and regularization terms. We demonstrate our algorithm's improvement over previous methods using simulated and real-world data, with several different elevation apertures.

In future work, more extensive evaluation of the proposed method in comparison to OGM [33, 34] and space carving [5] ought to be carried out with a wide variety of underwater structures. A drift-free SLAM method such as [35] should be utilized to increase the accuracy of sensor pose estimates and to allow for longer, richer datasets to be collected. The proposed method could be adapted to attempt to explicitly account for occlusions, for example by incorporating a non-convex prior as in [12].

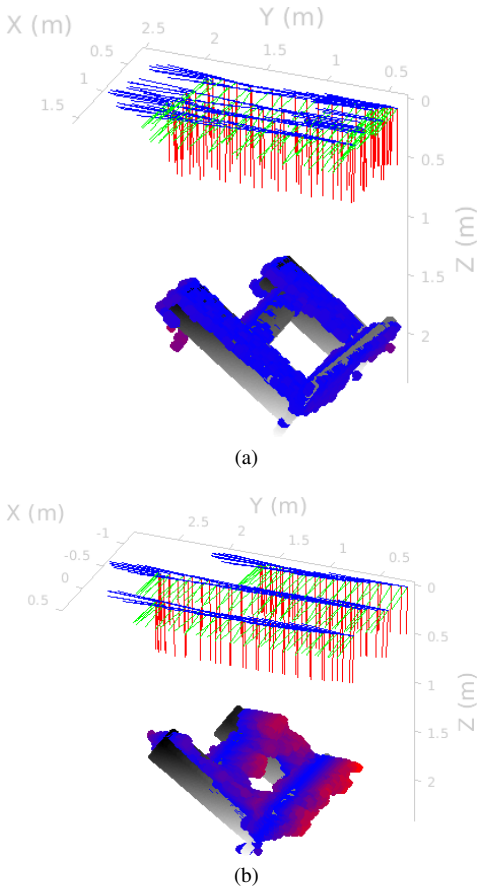


Fig. 6: Sample reconstruction of the imaged structure on (a) $1°$ and (b) $14°$ elevation aperture test tank datasets. Reconstructed points are colored blue to red with increasing error and are overlaid on a gray ground-truth pointcloud. Downsampled sensor poses are shown with the red, green, and blue coordinates axes corresponding to the sensor's x, y, and z axes, respectively.

References

[1] B. Ahn, A. Dave, A. Veeraraghavan, I. Gkioulekas, and A. C. Sankaranarayanan, "Convolutional approximations to the general non-line-of-sight imaging operator," in *Intl. Conf. on Computer Vision (ICCV)*, 2019.

[2] M. D. Aykin and S. Negahdaripour, "Forward-look 2-D sonar image formation and 3-D reconstruction," in *Proc. of the IEEE/MTS OCEANS Conf. and Exhibition*, 2013, pp. 1–10.

[3] ——, "On 3-D target reconstruction from multiple 2-D forward-scan sonar views," in *Proc. of the IEEE/MTS OCEANS Conf. and Exhibition*, May 2015, pp. 1949–1958.

[4] ——, "Modeling 2-D lens-based forward-scan sonar imagery for targets with diffuse reflectance," *IEEE J. of Oceanic Engineering*, vol. 41, no. 3, pp. 569–582, 2016.

[5] ——, "Three-dimensional target reconstruction from multiple 2-d forward-scan sonar views by space carving," *IEEE J. of Oceanic Engineering*, vol. 42, no. 3, pp. 574–589, 2016.

[6] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[7] M. Buttafava, J. Zeman, A. Tosi, K. Eliceiri, and A. Velten, "Non-line-of-sight imaging using a time-gated single photon avalanche diode," *Optics express*, vol. 23, no. 16, pp. 20 997–21 011, 2015.

[8] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 877–905, 2008.

[9] A. Elfes, "Using occupancy grids for mobile robot perception and navigation," *Computer*, vol. 22, no. 6, pp. 46–57, 1989.

[10] T. Guerneve and Y. Petillot, "Underwater 3d reconstruction using blueview imaging sonar," in *Proc. of the IEEE/MTS OCEANS Conf. and Exhibition*, 2015, pp. 1–7.

[11] T. Guerneve, K. Subr, and Y. Petillot, "Three-dimensional reconstruction of underwater objects using wide-aperture imaging sonar," *J. of Field Robotics*, vol. 35, no. 6, pp. 890–905, 2018.

[12] F. Heide, L. Xiao, W. Heidrich, and M. B. Hullin, "Diffuse mirrors: 3D reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors," in *Proc. IEEE Intl. Conf. Computer Vision and Pattern Recognition*, 2014, pp. 3222–3229.

[13] F. Heide, M. O'Toole, K. Zang, D. B. Lindell, S. Diamond, and G. Wetzstein, "Non-line-of-sight imaging with partial occluders and surface normals," *ACM Transactions on Graphics*, vol. 38, no. 3, p. 22,

2019.

[14] A. Hinduja, B.-J. Ho, and M. Kaess, "Degeneracy-aware factors with applications to underwater SLAM," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2019.

[15] B.-J. Ho, P. Sodhi, P. Teixeira, M. Hsiao, T. Kusnur, and M. Kaess, "Virtual occupancy grid map for submap-based pose graph SLAM and planning in 3D environments," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2018, pp. 2175–2182.

[16] T. A. Huang and M. Kaess, "Towards acoustic structure from motion for imaging sonar," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, Oct. 2015, pp. 758–765.

[17] T. Huang, "Acoustic structure from motion," Master's thesis, Carnegie Mellon University, Pittsburgh, PA, May 2016.

[18] H. Johannsson, M. Kaess, B. Englot, F. Hover, and J. Leonard, "Imaging sonar-aided navigation for autonomous underwater harbor surveillance," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2010, pp. 4396–4403.

[19] B. Kim and S.-C. Yu, "Imaging sonar based real-time underwater object detection utilizing AdaBoost method," in *Proc. of Intl. Symp. of Underwater Technology*, 2017, pp. 1–5.

[20] J. Kim and S.-C. Yu, "Convolutional neural network-based real-time ROV detection using forward-looking sonar image," in *IEEE/OES Autonomous Underwater Vehicle (AUV) Symposium*, 2016, pp. 396–400.

[21] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "A method for large-scale ℓ_1-regularized least squares," *IEEE J. on Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 606–617, 2007.

[22] D. Langer and M. Hebert, "Building qualitative elevation maps from side scan sonar data for autonomous underwater navigation," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 1991, pp. 2478–2483.

[23] S. Negahdaripour, "On 3-D motion estimation from feature tracks in 2-D FS sonar video," *IEEE Trans. Robotics*, vol. 29, no. 4, pp. 1016–1030, Aug. 2013.

[24] ——, "Application of forward-scan sonar stereo for 3-D scene reconstruction," *IEEE J. of Oceanic Engineering*, 2018.

[25] S. Negahdaripour, V. M. Milenkovic, N. Salarieh, and M. Mirzargar, "Refining 3-D object models constructed from multiple FS sonar images by space carving," in *Proc. of the IEEE/MTS OCEANS Conf. and Exhibition*, 2017, pp. 1–9.

[26] M. O'Toole, D. B. Lindell, and G. Wetzstein, "Confocal non-line-of-sight imaging based on the light-cone transform," *Nature*, vol. 555, no. 7696, p. 338, 2018.

[27] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *Proc. IEEE Intl. Conf. Computer Vision and Pattern Recognition*, vol. 1, 2006, pp. 519–528.

[28] Y. S. Shin, Y. Lee, H. T. Choi, and A. Kim, "Bundle adjustment from sonar images and SLAM application for seafloor mapping," in *Proc. of the IEEE/MTS OCEANS Conf. and Exhibition*, Oct. 2015, pp. 1–6.

[29] P. Sodhi, B.-J. Ho, and M. Kaess, "Online and consistent occupancy grid mapping for planning in unknown environments," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2019.

[30] P. Teixeira, M. Kaess, F. Hover, and J. Leonard, "Underwater inspection using sonar-based volumetric submaps," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, Oct. 2016, pp. 4288–4295.

[31] A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. G. Bawendi, and R. Raskar, "Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging," *Nature Communications*, vol. 3, p. 745, 2012.

[32] J. Wang, T. Shan, and B. Englot, "Underwater terrain reconstruction from forward-looking sonar imagery," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2019, pp. 3471–3477.

[33] Y. Wang, Y. Ji, H. Woo, Y. Tamura, A. Yamashita, and A. Hajime, "3D occupancy mapping framework based on acoustic camera in underwater environment," *IFAC-PapersOnLine*, vol. 51, no. 22, pp. 324–330, 2018.

[34] Y. Wang, Y. Ji, H. Woo, Y. Tamura, A. Yamashita, and H. Asama, "Three-dimensional underwater environment reconstruction with graph optimization using acoustic camera," in *IEEE/SICE Intl. Symp. on System Integration (SII)*, 2019, pp. 28–33.

[35] E. Westman and M. Kaess, "Underwater apriltag SLAM and extrinsics calibration for AUV localization," Robotics Institute, Carnegie Mellon University, Tech. Rep. CMU-RI-TR-18-43, Sep. 2018.

[36] ——, "Wide aperture imaging sonar reconstruction using generative models," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2019.

[37] ——, "Degeneracy-aware imaging sonar simultaneous localization and mapping," *IEEE J. of Oceanic Engineering*, 2019.

[38] E. Westman, A. Hinduja, and M. Kaess, "Feature-based SLAM for imaging sonar with under-constrained landmarks," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, May 2018, pp. 3629–3636.

[39] Y. Yang and G. Huang, "Acoustic-inertial underwater navigation," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, May 2017, pp. 4927–4933.

[40] B. Zerr and B. Stage, "Three-dimensional reconstruction of underwater objects from a sequence of sonar images," in *Proc. of the Intl. Conf. on Image Processing*, vol. 3, 1996, pp. 927–930.