Dense, Sonar-based Reconstruction of Underwater Scenes

Pedro V. Teixeira*, Dehann Fourie*, Michael Kaess†, and John J. Leonard*

***Abstract*— Typically, the reconstruction problem is addressed in three independent steps: first, sensor processing techniques are used to filter and segment sensor data as required by the front end. Second, the front end builds the factor graph for the problem to obtain an accurate estimate of the robot's full trajectory. Finally, the end product is obtained by further processing of sensor data, now re-projected from the optimized trajectory. In this paper we present an approach to model the reconstruction problem in a way that unifies the aforementioned problems under a single framework for a particular application: sonar-based inspection of underwater structures. This is achieved by formulating both the sonar segmentation and point cloud reconstruction problems as factor graphs, in tandem with the SLAM problem. We provide experimental results using data from a ship hull inspection test.**

I. Introduction

THREE-dimensional maps of underwater scenes are critical to—or the desired end product of—many applications over a spectrum of spatial scales. Examples range from microbathymetry and subsea inspection to hydrographic surveys of coastlines. Depending on the end use, maps will have different levels of required accuracy in the positions of the features they capture: the *IHO Standards for Hydrographic Surveys*, for instance, require a maximum horizontal uncertainty of $2m$ at a 95% confidence level [23]. Maps used for proximity navigation around sub-sea infrastructure are likely to have stricter requirements, as both the platform and features of interest are significantly smaller. The accuracy of a mapping platform depends mainly on the individual accuracies of: *(i)* its pose estimate in some global frame, *(ii)* the estimates of offsets between mapping sensors and platform, and *(iii)* the accuracy of the mapping sensor measurements. Typically, surface-based surveying platforms will employ highly accurate positioning sensors—*e.g.* a combination of differential Global Navigation Satellite System (GNSS) receiver with an accurate Attitude and Heading Reference System (AHRS)—to instrument the pose of a mapping sensor such as a multibeam sonar. Surveying is performed after a calibration of sensor offsets (usually through a set of dedicated maneuvers) and the data is finally fed to a post-processing tool to optimize. For underwater platforms, such as autonomous underwater vehicles, the use of absolute positioning systems is only possible when the survey area is small ($\mathcal{O}$(1km)) and, save for a few exceptions, the accuracy of these systems is significantly lower than that of differential GNSS. This performance reduction shifts the accuracy burden to the Inertial Navigation System (INS) and/or the position estimation framework, often necessitating the use of techniques such as Simultaneous Localization and Mapping (SLAM), as most INS will incur in drift over time.

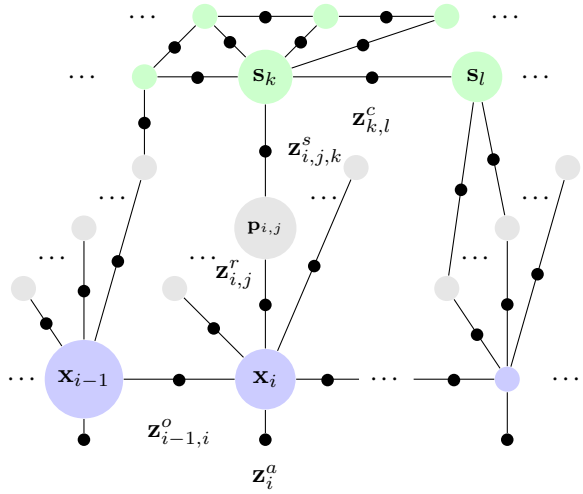


Fig. 1. Factor graph model for the reconstruction problem. Each vehicle pose node $\mathbf{x}_i$, in blue, is connected to the previous pose by a relative odometry measurement $\mathbf{z}^o_{i-1,i}$ (eq. 8). The node $\mathbf{x}_i$ is also constrained by the absolute depth, pitch, and roll measurement $\mathbf{z}^a_i$ (eq. 5). The range measurement $\mathbf{z}^r_{i,j}$ (eq. 3) connects the vehicle pose $\mathbf{x}_i$ to the corresponding scatterer position, $\mathbf{p}_j$, represented by the gray node. A scatterer may be associated with a surface element s_k, in which case it is constrained by the measurement $\mathbf{z}^s_{j,k}$ (eq. 10). Finally, adjacent surface elements may be constrained by a smoothness measurement, $\mathbf{z}^s_{k,l}$ (eq. 10). For clarity purposes, only one of each type of factor has been labeled in this figure.

In both surface and underwater survey platforms, there are three tasks that take place (often in real-time for autonomous platforms): (1) sensor processing, (2) pose estimation, and (3) 3D reconstruction. Their purpose is as follows: sensor processing aims at producing a set of measurements from the mapping sensor(s) that can be used to produce a map (and potentially by the pose estimation task as well, particularly in the context of SLAM); pose estimation tries to obtain accurate estimates of the platform pose over the entire survey trajectory; finally, 3D reconstruction uses the outputs from the previous two tasks to produce a map that is suitable to one or more end uses (e.g. navigation or inspection). As this description hints at, in many applications these tasks happen in sequence, with the output of one process being fed to the next with little to no feedback, potentially discarding information from the preceding steps that could produce a

This work was supported in part by the Office of Naval Research (under grant N00014-16-1-2103) and the Schlumberger Technology Corporation, which we gratefully acknowledge.

*Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 32 Vassar St, Cambridge, MA 02139, USA {pvt, dfourie, jleonard}@mit.edu

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

more accurate solution.

Unlike the sonar range measurement models found in many surface platforms [29, Sec. 6.3], where the sensor outputs a noisy range value, in most underwater mapping applications range estimates must be determined from single- or multi-beam sonar data on echo strength over range, often by taking the first or strongest return above a threshold. This is commonly preceded by the use of standard image processing operations as a pre-processing step to improve measurement accuracy [13], [3], [14]. To address the effects of outliers typically found in sonar data (caused by acoustic phenomena such as noise, reverberation and multi-path propagation) some of the proposed methods then look at the agreement between the range measurements in consecutive beams [18]. Another group of techniques makes this relationship explicit by modeling the problem of segmenting the *full* sonar image into free and occupied space using graphical models, such as Markov Random Fields [1], [28]. These *dense* methods, while more accurate, are computationally more intensive, and arguably less efficient for the purpose of obtaining a single range measurement per beam.

Considerably less common in the underwater domain is the use of scene models to aid sensor processing: knowing that the range measurements correspond to points along a smooth surface, for instance, can help with the segmentation process. Such approaches are commonplace in reconstruction applications, where noisy range data is associated with some form of surface representation, often based on geometric primitives such as *planes* [30], [7], *splines* [12] or *surfels* [33]. When not assumed fixed, the pose of the range sensor (or equivalently, that of the reconstructed object) is not usually measured; instead, sensor motion (*egomotion*) is estimated through use of Iterative Closest Point (ICP) variants from sequential range scans [33], [9]. Unfortunately, the combination of the geometry of multibeam sonars with platform motion frequently precludes the use of this family of algorithms—as platforms move perpendicular to the scanning plane to maximize coverage rate, overlap between consecutive images is eliminated. In cases where there is overlap due to in-plane motion, ICP can only provide partial (in-plane) constraints [14], and these tend to be poorly informative in the case of small fields of view. Thus, out-of plane motion generally leads to the use of submap-based techniques, where sensor measurements and odometry are accumulated over a short time frame to produce a "virtual" sensor measurement that can be registered with previous ones and produce a relative motion estimate [18], [27], [32].

While some reconstruction methods assume drift-free sensor trajectories, avoiding the *loop closure* problem [9], others formulate it as a full SLAM problem, estimating both sensor pose and primitive location and parameters. *Planar* SLAM is a prime example of the latter, in which the proposed methods take advantage of the ubiquity of planar features in man-made environments to concurrently use as landmarks and mitigate the effect of noise in the range measurements. Most approaches estimate both the sensor pose and the parameters of planes identified from two- and three-dimensional range sensor data [30], [7] which, while noisy, is quite dense when compared to typical sonar measurements (a notable exception is the use of very sparse range data to track planar features and derive navigation constraints [16]). One particularly relevant set of techniques uses a similar approach to refine the output of a SLAM system [20], [19]: modeling range measurements as surfels, the method optimizes both sensor pose and surfel parameters. This optimization is performed iteratively: once range measurements are approximated by surfels, pairwise correspondences are then determined; once new pose and surfel estimates are available, the graph is re-built and the process continues.

Drawing inspiration from these methods, the aim of this paper is then to address some of the limitations in prior work by the authors [27], [28] by formulating the reconstruction problem in a manner that allows for concurrent modeling and estimation of sensor pose, measurements, and model parameters in a single, unified framework. We accomplish this by formulating each of the three problems using the language of factor graphs. Owing to this modeling choice, we consider the proposed technique to be *dense*, as every valid range measurement is explicitly modeled as variable in the estimation problem.

After a more formal problem description in section II, we model sonar range measurements in section III, and tie those with a standard pose estimation formulation in section IV. Section V addresses the choice of surface representation, the associated measurement model, and the algorithms required for data association between the two. In section VI we present some results from the use of our proposed technique on a small segment of data taken from a larger inspection data set. Finally, in section VII we conclude with some remarks on the strengths and weaknesses of the technique, and point at promising directions for future work.

II. Problem Statement

Our problem can be described as that of estimating the position $\mathbf{p}$ of acoustic returns, as well as the pose $\mathbf{x}$ of the sensor itself. Provided an adequate scene model $\mathcal{S}$, we would also like to estimate its parameters concurrently. The inputs to our estimation problem comprise multibeam sonar scans and navigation data, the latter provided either as raw sensor measurements (e.g. from a combination of IMU, DVL, and pressure sensor), or as odometry estimates $\hat{\mathbf{x}}_i$ from an external navigation system. Multibeam sonar scans contain a set of measured intensity values $y_k(r)$ defined over the detection range $[r_{min}, r_{max}]$, where k indexes the beam in the scan ($k \in \{1, \ldots, B\}$). The choice of a scene model is addressed later in the text (Sec. V).

III. Sonar Segmentation

The purpose of segmenting a sonar scan is to determine, given $y_k(r)$, a belief over the range to the scattering object or surface—$p_k(r|y)$—for each beam in the scan. In the ideal case, determining the range to the object amounts to detecting the presence and location (r_k) of a transmitted signal $u(r)$ in the received signal $y(r)$, subject to additive

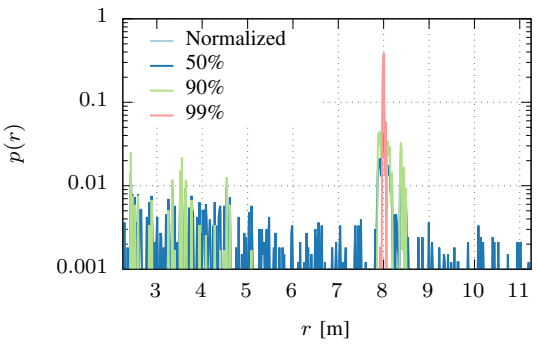


Fig. 2. Normalized empirical distributions obtained from the a typical single-beam echo intensity measurement. The different curves show the effect of removing the lowest 50, 90, and 99% quantiles.

noise and linear medium response: $y_k(r) = u(r-r_k)+n_k(r)$. This is also known as *matched filtering*, and under these assumptions $p(r|y)$ can be recovered given the received and transmitted signals [35]:

$$p(r|y) = z^{-1}p(r)\exp\left(q(r)\right) \tag{1}$$

where z is a normalizing constant, $q(r)$ is the correlation of the transmitted pulse with the received signal,

$$q(r) = \frac{2}{N_0}\int_{r_{min}}^{r_{max}} y(s)u(s-r)ds \tag{2}$$

and N_0 is the energy in the noise signal (assumed white). If the prior belief on r is uniform, then the MAP estimate of r is simply the maximum of $q(r)$.

In other situations, however, we may have no knowledge of the transmitted pulse, and we are left with the problem of recovering $p(y|r)$ from $y(r)$ alone. Here, we assume $p(r|y) \propto y(r)$ will not be Gaussian, or equivalently $u(r) = \delta(r)$ (impulse at r). In fact, $p(y|r)$ may contain more than one mode due to reverberation, multi-path propagation, and other acoustic phenomena, as illustrated in figure 2 with normalized $y(r)$. Figure 2 also highlights the multi-modal nature of the measurement by removing the lowest 50, 90, and 99% quantiles of $y(r)$.

So far, we have concerned ourselves with modeling the problem along the sonar beam's main axis (i.e. *range*); to pinpoint the position of the received echo in space, we must also consider the remaining two directions[1]: azimuth (α) and elevation (β). The uncertainty associated with these axes stems from the fact that the beam has a non-zero width. We assume azimuth and elevation to be independent and treat $p(\alpha)$ and $p(\beta)$ as static, approximated by a bivariate normal with σ_α and σ_β equal to the width of the main lobe in the respective direction (usually on the order of 1°). The measurement model for the sonar is

$$\begin{aligned}\mathbf{z}_i^r &= h(\mathbf{x}_j^s, \mathbf{p}_k) + \mathbf{v}_i^r \\ &= \begin{bmatrix} \|\mathbf{x}_j^s - \mathbf{p}_k\|_2 \\ \arctan(\frac{p_y - y}{p_x - x}) \\ \arccos(p_z - z, r) \end{bmatrix} + \mathbf{v}_i^r \\ &= [r\ \alpha\ \beta]^T + \mathbf{v}_i^r \end{aligned} \tag{3}$$

[1]Here we employ a spherical coordinate frame, as is typically the case when working with sonar systems.

where the covariance matrix $\Sigma = diag([\sigma_r^2\ \sigma_\alpha^2\ \sigma_\beta^2]^T)$ is obtained from uncertainty in the range estimate (σ_r) and main lobe width in the azimuth (σ_α) and elevation (σ_β) directions. $\mathbf{x}_j^s$ and $\mathbf{p}_k$ are the sensor pose and return position (respectively), expressed in the world frame.

Given the MAP estimate for the return position ${}^S\mathbf{p}_k$ (or, equivalently, estimates for range, azimuth, and elevation—r, α, and β), the return can then be registered in the world frame:

$${}^W\tilde{\mathbf{p}}_k = {}^W_P\hat{T}\ {}^P_S\hat{T}_i\,{}^S\tilde{\mathbf{p}}_k \tag{4}$$

where *W*, *P*, and *S* denote the *world*, *platform*, and *sensor* frames, respectively, and $\tilde{\mathbf{p}}$ denotes the homogeneous representation of point $\mathbf{p}$[2]. Figure 4a shows the point cloud obtained through registration of sonar returns from odometry-based estimates of ${}^W_P\hat{T}$.

IV. Pose Estimation

To spatially register sonar returns (eq. 4) we require estimates of both sensor and platform poses to be available—these must be obtained from measurements of the on-board navigation sensors. In this section, we describe the formulation of the navigation part of the estimation problem, assuming a typical navigation payload comprising a Doppler velocity log (DVL), inertial measurement unit (IMU), and pressure sensor [10]. Absolute and relative measurements are captured by unary and binary constraints in our factor graph model, as illustrated in figure 1.

A. Absolute Measurements—Depth, Pitch, and Roll

Depth estimates can be obtained from pressure measurements through the use of seawater models [4]. Pitch and roll measurements, in turn, are available directly from the platform's AHRS. The measurement model is then

$$\begin{aligned}\mathbf{z}_i^a &= h(\mathbf{x}_i) + \mathbf{v}_i^a \\ &= [z\ \theta\ \phi]^T + \mathbf{v}_i^a \end{aligned} \tag{5}$$

The measurement covariance is $\Sigma = diag([\sigma_z^2\ \sigma_\theta^2\ \sigma_\phi^2])$, with typical values of $0.1\ m$ for depth[3], and 0.1° for pitch and roll [10].

B. Relative Measurements—Horizontal Odometry

Horizontal odometry measurements are obtained through integration of the platform velocity, ${}^P\mathbf{u} = [u\ v\ w]^T$ [34]:

$$\begin{aligned}\begin{bmatrix}\delta x \\ \delta y \\ \delta z\end{bmatrix}_{i,i+1} &= \int_{t_i}^{t_{i+1}} {}^W_P R(t)\,{}^P\mathbf{u}(\mathbf{t})dt \\ &\approx {}^W_P R_i\,{}^P\mathbf{u}_i\delta t \end{aligned} \tag{6}$$

and of the z-component of the angular velocity in the platform frame (small-angle approximation):

$$\begin{aligned}\delta\psi &\approx \int_{t_i}^{t_{i+1}} {}^P\omega_z(t)dt \\ &\approx {}^P\omega_{z,i}\delta t \end{aligned} \tag{7}$$

[2]The homogenous representation of point $\mathbf{p}$ is the vector $\tilde{\mathbf{p}} = [\mathbf{p}^T\ 1]^T$.
[3]Assuming shallow depth ($\leq 100\ m$), and σ around 0.1% of full scale.

where δt is the sampling period. The horizontal odometry measurement model is then

$$\begin{aligned}\mathbf{z}_i^o &= h(\mathbf{x}_i, \mathbf{x}_{i+1}) + \mathbf{v}_i^o \\ &= [\delta x \ \delta y \ \delta\psi]^T + \mathbf{v}_i^o\end{aligned} \tag{8}$$

Due to noise in the measurements $\mathbf{u}$, ω, as well as bias in the latter, the covariance associated with $\mathbf{z}_i^o$ will grow with time, modeled by $\Sigma = t{\cdot}diag([\sigma_x^2 \ \sigma_y^2 \ \sigma_\psi^2 \ ])$. Note that the equations above necessitate a prior transformation of the linear and angular velocity vectors from their respective frames to the platform frame, thus requiring proper calibration so that ${}^P_{DVL}T$ and ${}^P_{IMU}T$ can be accurately determined [11].

C. Sensor Offset

Like the DVL and IMU, the mapping sensor payload also requires calibration to determine ${}^P_S T$, as lever arm effects can introduce registration errors in the order of tens of centimeters when mapping at a long range. To address the issue, we model the offset between sensor and platform as part of the estimation problem: with every new sonar measurement a new measurement is added:

$$\begin{aligned}\mathbf{z}_i^c &= h(\mathbf{x}_i^p, \mathbf{x}_i^s, c) + v_i^c \\ &= {}^P_S\hat{T} \ {}^W_S\hat{T}_i^{-1} \ {}^W_P\hat{T}_i + v_i^c\end{aligned} \tag{9}$$

This measurement model describes what is essentially a "consistency check" between the platform and sensor poses and the sensor offset; if all three estimates are correct, the measurement should yield the identity transformation.

V. Reconstruction

Thus far, we have considered a point-based representation of the scene, where the position of the sonar returns in the world frame can be obtained by projecting the most likely range value (eq. 4) from the sensor to the world frame. What this formulation has not yet captured, however, is that these points are, in fact, noisy samples of some object surface. Given a surface representation, we can model this as a constraint between the surface and the point sample.

Common candidates for discrete representation include simple geometric primitives, such as lines, planes, and *surfels* [15], as well as parametric surfaces [12]. As mentioned in Section I, the choice of a particular primitive (or set of primitives) is tied to the characteristics of the scene: while planes tend to be a good fit to man-made environment, the same does not hold true for less structured environments, such as underwater scenes. Other methods eschew geometric primitives in favor of a (minimal) set of features derived from the actual scene [21], but approach the problem from a data compression perspective, taking the representation as the fixed output of a mapping system.

For these reasons, we use the small-scale, spatially bounded version of planes—*surfels*—as the discrete representation of choice. Modeled by an origin $\mathbf{o}$ and a surface normal $\hat{\mathbf{n}}$ of unit length, surfels are also characterized by their spatial support r_S, which we consider a reconstruction parameter instead of part of the estimation problem. Thus, we implicitly make the assumption that, for the desired level of accuracy, the scene can be approximated by a piecewise planar set of primitives; in other words the scale of characteristic features in the scene is comparable to, or larger than, the spatial support r_S.

A. Surfel Correspondence

Given a point $\mathbf{p}_j$ and a *surfel* $\mathbf{s}_k$, we model a correspondence or assignment measurement as a combination of the *point-to-plane* (d_o) and *in-plane* (d_i) distances between the two, which can be written as

$$\begin{aligned}\mathbf{z}_{j,k}^s &= h(\mathbf{p}_j, \mathbf{s}_k) + \mathbf{v} \\ &= [d_o \ d_i]^T + \mathbf{v} \\ &= \begin{bmatrix}\sqrt{(\mathbf{n_k}^T(\mathbf{p_j} - \mathbf{o_k}))^2} \\ \sqrt{\|\mathbf{p_j} - \mathbf{o_k}\|^2 - d_o^2}\end{bmatrix} + \mathbf{v}\end{aligned} \tag{10}$$

If $\mathbf{p}_j$ is considered to be a sample of s_k, then the expected measurement value should be zero, with covariance of $\Sigma = diag([\sigma_o^2 \ \sigma_i^2])$, where σ_o should be closely related to the variance (width) of the sonar return chosen as the range measurement (Sec. III). While not strictly necessary, the in-plane distance component of the measurement model serves to address the degeneracy issues associated with over-parameterized plane representations [7]; for this reason, σ_i should not dominate over σ_r, and should in fact be proportional to the surfel support r_S.

B. Smoothness constraints

Another implicit assumption in the approximation of a surface with a set of surfels smaller than (or comparable to) the characteristic scale of its features, is that these should be somewhat evenly distributed, and that their orientation should vary smoothly. To model this potential smoothness constraint between neighboring surfels, we use a two-dimensional measurement comprising the pairwise point to plane distances between one surfel's origin and the other's plane:

$$\begin{aligned}\mathbf{z}_{j,k}^c &= h(\mathbf{s}_j, \mathbf{s}_k) + \mathbf{v}_c \\ &= \begin{bmatrix}\sqrt{(\mathbf{n_j}^T(\mathbf{o_k} - \mathbf{o_j}))^2} \\ \sqrt{(\mathbf{n_k}^T(\mathbf{o_j} - \mathbf{o_k}))^2}\end{bmatrix} + \mathbf{v}_c\end{aligned} \tag{11}$$

The noise model for smoothness constraints is governed by a single parameter, $\Sigma_c = \sigma_c^2 I_{2\times 2}$, controlling how tightly the constraints are enforced.

C. Segmentation

Having described the measurement models for both the sonar range measurement (eq. 3) and the associated scattering surface element (eq. 10), we now turn to the data association problem of how to pair points to surfels. Starting from a set of scans and associated odometry, which we use to spatially register the range measurements and obtain an initial point cloud (figure 4a), we would like to derive a segmented cloud comprising *(i)* the pairwise assignments between points and surfels, and *(ii)* the adjacency between these patches, which will inform the use of continuity constraints.

While there are several potential methods to address this necessity [25], *Voxel Cloud Connectivity Segmentation*

(VCCS) [17] stands out, as it fulfills both requirements. Still, due to the different clustering criteria in this application (explained later in the text), we opt for implementing an arguably simpler (but potentially less efficient) method. Furthermore, if real-time operation is desired, the method must support incremental execution, as new, unsegmented, sensor data arrives.

To facilitate the description of our incremental segmentation approach, we represent sonar returns as points $\mathbf{p} = [\mathbf{x}^T\ \hat{\mathbf{n}}^T\ l\ t]^T$, where $\mathbf{x} = [x\ y\ z]^T$ and $\hat{\mathbf{n}} = [n_x\ n_y\ n_z]^T$ are the point's location and (unit length) normal. l and t are the point's label and acquisition time. We denote as $\mathcal{P}_l$ the set of points with label l (setting $l = 0$ for unlabeled points), and $\mathcal{N}_{\mathbf{p}}(\mathcal{P})$ as the points in $\mathcal{P}$ that neighbor point $\mathbf{p}$[4]. Similarly, we define the set of *seed points* as $\mathcal{S}$, containing exactly one point per label. To keep track of adjacency between patches, we use the graph $\mathcal{G} = (\mathcal{S}, \mathcal{E})$, where $\mathcal{E}$ is the set of undirected edges (i, j) connecting patches i and j. Finally, we define a *comparison operator* $C(\mathbf{p}_i, \mathbf{p}_j)$, which equals true when $\mathbf{p}_i, \mathbf{p}_j$ are *similar* [31]. This definition requires that $\mathbf{p}_i$ and $\mathbf{p}_j$ are expressed in a common reference frame; this is accomplished by registering each of the points in the world frame using the associated (odometry-based) platform pose estimate.

The first step in the algorithm is to generate new seed points from the subset of unlabeled points, $\mathcal{P}_0$. This is accomplished by subsampling the point set through the use of a voxel grid with a resolution of r_{seed}, which we set to be equal to twice the spatial support r_S. If, in addition, each voxel is required to contain a minimum number of points, some level of outlier rejection can be achieved at the expense of not segmenting sparsely mapped areas. Once new seed points s' have been added to the seed point set $\mathcal{S}$, the adjacency graph $\mathcal{G}$ is updated by looking, for each new seed point, for similar seed points in its neighborhood: if $C(\mathbf{s}', \mathbf{s}),\ \mathbf{s} \in \mathcal{N}_{s'}(\mathcal{S})$, then the two are considered adjacent and the edge $(\mathbf{s}', \mathbf{s})$ is added to $\mathcal{E}$. Similarly, pairing points with seed points is also performed using a greedy approach: for each unsegmented point $\mathbf{p} \in \mathcal{P}_0$, we iterate from the closest to farthest seed point within a search radius, and set its label to that of the first similar seed point, i.e. $l(\mathbf{p}_j) = l(\mathbf{s}_k)$ if $C(\mathbf{p}_j, \mathbf{s}_k)$ is true. Since we are trying to approximate the scene by a set of small planar patches, two points should be considered similar if their normals are aligned, *i.e.*, if $\hat{\mathbf{n}}_j \cdot \hat{\mathbf{n}}_k \geq \cos(\alpha_{max})$, where α_{max} is the maximum allowable angle between the two normals. Due to the issue of drift in the sensor position estimate, it may happen that the odometry-based estimates of two separate points $\mathbf{p}_j$ and $\mathbf{p}_k$ end up in close proximity. To avoid this scenario, two points should only be considered similar if the time span between their acquisitions is short - in other words, they must be close in space *and* time[5]. Thus, we define the comparison operator as:

$$C(\mathbf{p}_i, \mathbf{p}_j) = (|t_i - t_j| < \delta t_{max}) \wedge (\hat{\mathbf{n}}_i \cdot \hat{\mathbf{n}}_j \geq \cos \alpha_{max}) \tag{12}$$

Given an updated point set $\mathcal{P}$ (with new unlabeled points), the three steps above are repeated until there are no new pairings, at which point the potentially non-exhaustively labeled point set $\mathcal{P}$ and graph $\mathcal{G}$ are used to update the factor graph through the addition of new patch nodes, point and patch associations for newly paired points (eq. 10), and smoothness constraints (eq. 11) for adjacent surfels. To keep with the definition of surfel, only patches with curvature below a maximum value are added to the factor graph. Finally, it is also worth noting that the need for iterative segmentation within each incremental update is driven by the seed point generation mechanism: if a voxel contains two sets of dissimilar points, acquired around t_i and t_k, just one of these sets will be segmented after one step, as only one seed point will have been chosen from each voxel per iteration. By iterating seed point generation, we minimize the number of unlabeled points.

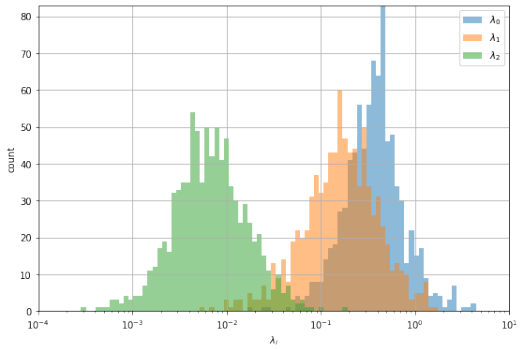


Fig. 3. Principal component distribution for valid patches in the dataset—these are planar ($\lambda_2 / \sum \lambda \ll 1$), and nearly circular ($\lambda_0 \approx \lambda_1$).

VI. Experimental Results

A. Dataset

The data used for experimental evaluation of the proposed method is a segment of a ship hull inspection test with the Hovering Autonomous Underwater Vehicle (HAUV) [6], comprising 750 scans from a DIDSON multibeam sonar [2] acquired over a span of approximately two minutes.

B. Parameters

Based on the sonar properties [2], we set $\sigma_\alpha = 0.3°$ and $\sigma_\beta = 1.0°$, and use a conservative $\sigma_r = 0.05m$ for range. For odometry measurements, we let $\Sigma_a = diag([0.1^2\ \ (1°)^2\ \ (1°)^2]^T)$ and $\Sigma_o = \delta_i t \cdot diag(1 \times 10^{-3}[3\ 3\ 1]^T)$ for absolute and relative odometry measurements, respectively, where $\delta t_i = t_i - t_{i-1}$. For this particular segment, the platform moves laterally (perpendicular to the sonar plane) with $|v| \approx 0.2\ m.s^{-1}$, so we let $r_S = 0.15\ m$ and $\delta t_{max} = 2\ s$, and require a minimum of 10 points to generate a seed.

[4] Typically, the neighborhood of $\mathbf{q}$ in $\mathcal{P}$ comprises all points $\mathbf{p}$ in $\mathcal{P}$ within a distance r from $\mathbf{q}$: $\mathcal{N}_{\mathbf{q}}(P) = \{\mathbf{p} \in \mathcal{P} : \|\mathbf{p} - \mathbf{q}\| < r\}$.

[5] This time difference criterion is used to determine the points in $\mathcal{N}_{\mathbf{p}}(\mathcal{P})$ from which the normal $\hat{\mathbf{n}}$ and curvature $c = \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}$ for $\mathbf{p}$ are estimated.

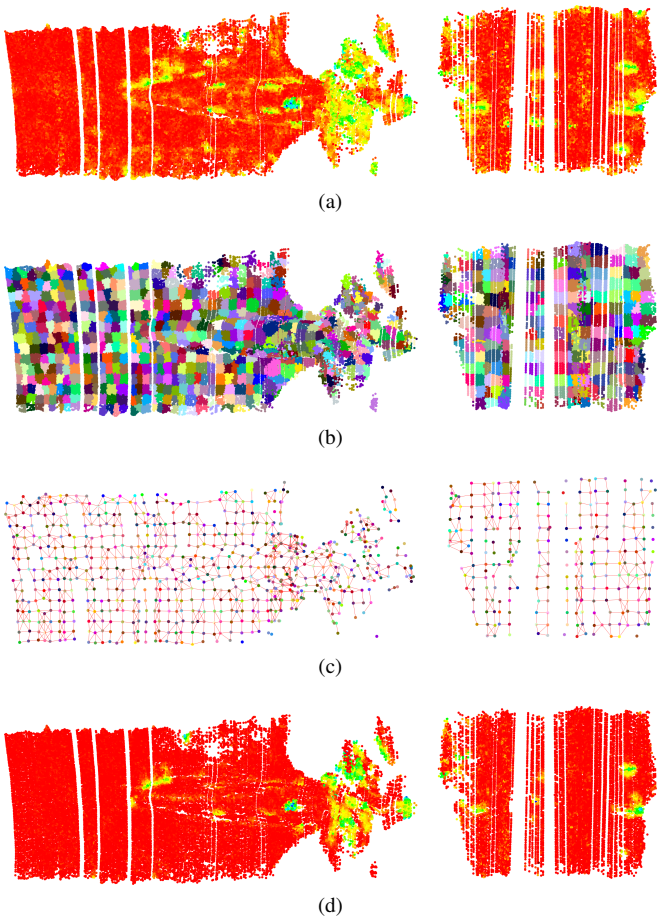
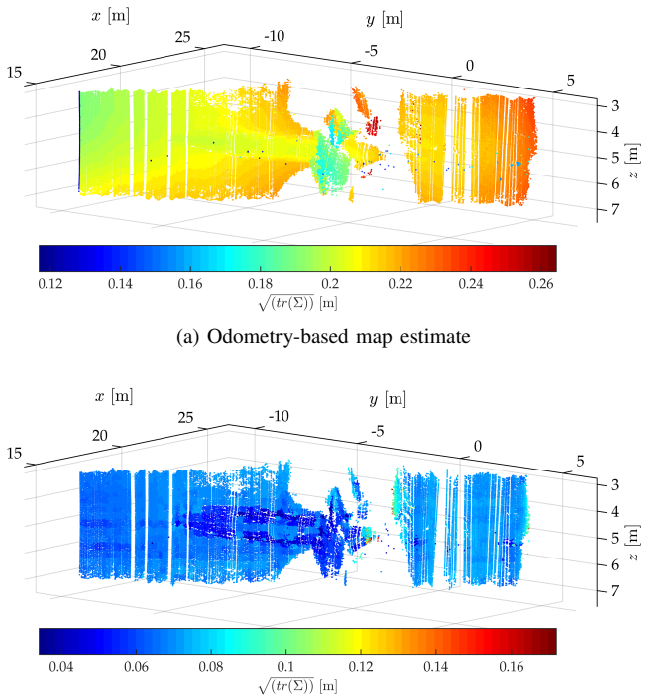


Fig. 4. Segmentation output: input point cloud (a), adjacency graph $\mathcal{G} = (\mathcal{S}, \mathcal{E})$ (b), segmentation output (c), and output point cloud (d). Input and output point clouds (a,d) are colored by local curvature ($c = \frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3}$). The reconstruction, with a scale of $r = 0.15m$, generated 1014 seed points and 868 valid (planar) patches. Vertical gaps in the point cloud are caused by missing sonar measurements.



(a) Odometry-based map estimate

(b) SLAM-based map estimate

Fig. 5. Odometry- (top) and SLAM-based (bottom) map estimates, where each point is colored by the associated uncertainty—$\sqrt{tr(\Sigma)}$. The uncertainty is initially dominated by the vertical (z) component; as the vehicle moves (left to right), the horizontal uncertainty grows and dominates. The proposed method improves map accuracy by reducing uncertainty by a factor of 2-5.

C. Reconstruction Results

We implemented the methods described in sections III through V, leveraging iSAM [8] for factor graph optimization, and PCL [22] for point cloud processing and visualization. The results of the algorithm on the ship hull data set are shown in figures 3 through 5.

Out of the initial 51,018 points, the resulting reconstruction generated 1020 seed points, but only 868 valid patches and 2198 edges, shown in figures 4b and 4c, respectively. Only 896 points were left unsegmented—a loss of approximately 2%. Figure 3 shows the distribution of the principal components of valid patches, which are planar (and nearly circular), as assumed in Section V.

Figures 4a and 4d show the impact of the choice of r_S: while most of the surface roughness induced by range measurement noise has been removed, so have some of the small scale detail, such as the four zinc anodes ($< r_S$) forward of the propeller. The larger object ($\approx r_S$) between the propeller and anodes, however, is still noticeable. This is visible in figures 4a and 4d, where the local curvature of the output point cloud has been significantly reduced. The support parameter r_S governs the spatial accuracy of the reconstruction—a large value acts as a spatial low-pass filter. On the other hand, decreasing r_S can lead to poor segmentation performance in sparsely covered areas, as few to no seed points will be generated. Similarly, the parameters σ_o and σ_r govern the trade-off between reliance on the original sonar measurement versus the piecewise planar approximation.

VII. Conclusion and Future Work

The proposed method addresses the artificial separation between sensor processing, pose estimation, and model reconstruction in the scope of sonar-based mapping. Leveraging scene information to aid in sonar processing requires a shared model between the two tasks—we choose one commonly used for pose estimation: *factor graphs*. The surfel-based, piecewise planar approximation of a scene was proven to work experimentally, increasing map accuracy without loop closures, relying instead on scene and sensor models. Still, it requires that some attention be paid to certain parameters, namely, the spatial support/characteristic scale r_S, and the relative weight of range and surfel sample constraints, σ_r and σ_s. In particular, σ_r is likely to be predetermined by the sonar and scene properties, as the return signal will depend on them.

One of the limitations of the approach presented in this article is that it does not contemplate long-term loop closures, needed to mitigate the growing uncertainty in the pose estimate. This has since been addressed by leveraging the

surfel graph $\mathcal{S}$ to derive loop closures as sets of pairwise correspondences between surfels [26]. Ongoing work aims at improving these techniques by leveraging relevant methods to extract higher-level features from the surfel graph [24]. Another important area for future work is the scalability of the proposed approach: by modeling every valid range measurement, pose, and surfel, the size of factor graph will quickly grow to the point where real-time performance is not feasible for all but the simplest problems. To mitigate this growth in complexity, we plan on modifying the proposed approach to avoid explicitly modeling point variables—the dominant factor in problem dimensionality. Finally, future research could also aim at leveraging non-parametric methods [5] to relax the Gaussianity assumptions made in the sonar measurement model (eq. 3). Such methods would capture the inherently multi-modal distributions associated with a sonar measurement.

ACKNOWLEDGEMENTS

The authors would like to thank the reviewers for the detailed and helpful comments on an earlier version of this article.

REFERENCES

[1] D. A. Abraham and P. K. Willett, "Active sonar detection in shallow water using the Page test," *IEEE J. Oceanic Eng.*, vol. 27, no. 1, pp. 35–46, Jan. 2002.

[2] E. Belcher, W. Hanot, and J. Burch, "Dual-frequency identification sonar (DIDSON)," in *Proceedings of the 2002 Interntional Symposium on Underwater Technology*, 2002, pp. 187–192.

[3] A. Burguera, G. Oliver, and Y. González, "Range extraction from underwater imaging sonar data," in *IEEE Conf. Emerging Technologies and Factory Automation (EFTA)*, Sep. 2010, pp. 1–4.

[4] N. P. Fofonoff and R. C. Millard Jr, "Algorithms for the computation of fundamental properties of seawater," ser. Unesco Technical Papers in Marine Science. Unesco, 1983, no. 44.

[5] D. Fourie, "Multi-modal and inertial sensor solutions to navigation-type factor graphs," Ph.D. dissertation, Massachusetts Inst. of Technology, Cambridge, MA, USA, Sep. 2017.

[6] 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, Jun. 2007.

[7] M. Kaess, "Simultaneous localization and mapping with infinite planes," in *IEEE Int. Conf. Robotics and Automation (ICRA)*, May 2015, pp. 4605–4611.

[8] M. Kaess, A. Ranganathan, and F. Dellaert, "iSAM: Incremental smoothing and mapping," *IEEE Trans. Robotics*, vol. 24, no. 6, pp. 1365–1378, Dec. 2008.

[9] M. Keller, D. Lefloch, M. Lambers, S. Izadi, T. Weyrich, and A. Kolb, "Real-time 3D reconstruction in dynamic scenes using point-based fusion," in *Int. Conf. 3D Vision*, Jun. 2013.

[10] J. C. Kinsey, R. M. Eustice, and L. L. Whitcomb, "A survey of underwater vehicle navigation: Recent advances and new challenges," in *IFAC Conf. Manoeuvering and Control of Marine Craft*, vol. 88, 2006.

[11] J. C. Kinsey and L. L. Whitcomb, "Towards in-situ calibration of gyro and doppler navigation sensors for precision underwater vehicle navigation," in *IEEE Int. Conf. Robotics and Automation (ICRA)*, vol. 4, 2002, pp. 4016–4023.

[12] G. Kurz and U. D. Hanebeck, "Dynamic surface reconstruction by recursive fusion of depth and position measurements," *Journal of Advances in Information Fusion*, vol. 9, no. 1, pp. 13–26, 2014.

[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 *IEEE OCEANS*, May 2009.

[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 *IEEE Int. Conf. Robotics and Biomimetics (ROBIO)*, Dec. 2012, pp. 538–544.

[15] F. Nardi, B. D. Corte, and G. Grisetti, "Unified representation and registration of heterogeneous sets of geometric primitives," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 625–632, Apr. 2019.

[16] P. Ozog and R. M. Eustice, "Real-time SLAM with piecewise-planar surface models and sparse 3D point clouds," in *IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, Nov. 2013, pp. 1042–1049.

[17] J. Papon, A. Abramov, M. Schoeler, and F. Wörgötter, "Voxel cloud connectivity segmentation - supervoxels for point clouds," in *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Portland, Oregon, Jun. 2013.

[18] C. Roman, "Self consistent bathymetric mapping from robotic vehicles in the deep ocean," Ph.D. dissertation, Massachusetts Inst. of Technology, Cambridge, MA, USA, 2005.

[19] M. Ruhnke, R. Kummerle, G. Grisetti, and W. Burgard, "Highly accurate maximum likelihood laser mapping by jointly optimizing laser points and robot poses," in *IEEE Int. Conf. Robotics and Automation (ICRA)*, May 2011, pp. 2812–2817.

[20] ——, "Range sensor based model construction by sparse surface adjustment," in *Advanced Robotics and its Social Impacts*, Oct. 2011, pp. 46–49.

[21] M. Ruhnke, B. Steder, G. Grisetti, and W. Burgard, *3D Environment Modeling Based on Surface Primitives*, Berlin, Heidelberg, 2012, pp. 281–300.

[22] R. B. Rusu and S. Cousins, "3D is here: Point Cloud Library (PCL)," in *IEEE Int. Conf. Robotics and Automation (ICRA)*, Shanghai, China, May 2011.

[23] I. S-44, "IHO standards for hydrographic surveys," International Hydrographic Organization, Special Publication 44, Feb. 2008.

[24] S. C. Stein, F. W. M. Schoeler, J. Papon, and T. Kulvicius, "Convexity based object partitioning for robot applications," in *IEEE Int. Conf. Robotics and Automation (ICRA)*, May 2014, pp. 3213–3220.

[25] D. Stutz, "Superpixel segmentation: An evaluation," in *Lecture Notes in Computer Science*. Springer International Publishing, 2015, pp. 555–562.

[26] P. V. Teixeira, "Dense, sonar-based reconstruction of underwater scenes," Ph.D. dissertation, Massachusetts Inst. of Technology, Cambrige, MA, USA, Sep. 2019.

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

[28] ——, "Multibeam data processing for underwater mapping," in *IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, Oct. 2018, pp. 1877–1884.

[29] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. MIT Press, 2005.

[30] A. J. B. Trevor, J. G. Rogers, and H. I. Christensen, "Planar surface SLAM with 3D and 2D sensors," in *IEEE Int. Conf. Robotics and Automation (ICRA)*, May 2012, pp. 3041–3048.

[31] A. J. B. Trevor, S. Gedikli, R. B. Rusu, and H. I. Christensen, "Efficient organized point cloud segmentation with connected components," in *IEEE Int. Conf. Robotics and Automation (ICRA)*, 2013, workshop on Semantic Perception Mapping and Exploration (SPME).

[32] M. A. VanMiddlesworth, M. Kaess, F. S. Hover, and J. J. Leonard, "Mapping 3D underwater environments with smoothed submaps," in *Int. Conf. Field and Service Robotics (FSR)*. Springer International Publishing, 2015, pp. 17–30.

[33] T. Weise, T. Wismer, B. Leibe, and L. V. Gool, "In-hand scanning with online loop closure," in *IEEE Int. Conf. Computer Vision (ICCV)*, Sep. 2009, pp. 1630–1637.

[34] L. Whitcomb, D. Yoerger, and H. Singh, "Advances in doppler-based navigation of underwater robotic vehicles," in *IEEE Int. Conf. Robotics and Automation (ICRA)*, vol. 1, 1999, pp. 399–406.

[35] P. M. Woodward, *Probability and Information Theory, with Applications to Radar*, ser. Electronics and Waves, D. W. Fry, Ed. Pergamon Press, 1953.