# Underwater Inspection using Sonar-based Volumetric Submaps

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

*Abstract*— **We propose a submap-based technique for mapping of underwater structures with complex geometries. Our approach relies on the use of probabilistic volumetric techniques to create submaps from multibeam sonar scans, as these offer increased outlier robustness. Special attention is paid to the problem of denoising/enhancing sonar data. Pairwise submap alignment constraints are used in a factor graph framework to correct for navigation drift and improve map accuracy. We provide experimental results obtained from the inspection of the running gear and bulbous bow of a 600-foot, Wright-class supply ship.**

## I. INTRODUCTION

Underwater mapping has a broad range of applications, including bathymetry, archaeological surveying, and the inspection of complex underwater structures such as subsea equipment or ship hulls. As a result of the harsh environmental conditions, mapping and inspection tasks are often performed by robots (either autonomous or remotely operated), equipped with optical and/or acoustic sensor payloads (e.g. cameras and sonars) typically capable of at least centimeter resolution. Still, the resolution of the resulting maps is usually lower, as the registration of sensor data is limited by the uncertainty of the navigation solution. As it grows over time, this uncertainty quickly exceeds the resolution of the sensors being used. While some applications—such as bathymetric surveys of relatively open area—may allow the use of absolute positioning systems that can limit this growth (e.g. LBL or USBL), this is rarely the case when dealing with structures such as ship hulls; in such situations, the complex acoustic environment significantly impacts the operation and performance of such systems due to effects such as multi-path or occultation.

Most of the sensors used in underwater mapping and inspection fall under the aforementioned categories—optical and acoustic. Optical systems are of particular interest due to their high resolution and low cost. Feature-based approaches for optical imagery have been used to successfully navigate and map a ship's hull [1]. Using features extracted from stereo camera pair imagery, a "dense"[1] smoothing technique is employed, allowing for tighter bounds to be placed on the navigation uncertainty. One of the main disadvantages of optical systems, however, is their sensitivity to underwater visibility conditions: with increased turbidity (common in harbors and near-shore environments), the range of optical sensors decreases, negatively impacting their coverage rate, if not preventing their operation altogether. This limitation has been addressed successfully in the context of ship hull inspection by utilizing an imaging sonar—again in a feature-based smoothing approach [2]. A common aspect to these techniques is that both target the smooth sections of the hull—the underlying assumption being that these are sufficiently smooth so that local planar approximations can be made, onto which features can then be registered. This is not a limitation of either technique, but rather, of the sensor used (imaging sonars, for instance, have an inherent elevation ambiguity). Geometrically complex scenes, such as a propeller, break this assumption and require us to use other sensors. While optical sensors are available that let us circumvent this limitation, such as ranging lasers based on either time-of-flight or structured light, they're often very costly and still subject to visibility issues, precluding their use in many man-made environments. Instead, their functionally equivalent acoustic counterparts, single- or multi-beam sonars, are often used for underwater mapping of more complex scenes.

Some of the earliest work on underwater mapping used both ship- and ROV-mounted multi-beam sonars to create bathymetric maps and inspect ship wrecks [3]. The sensor model utilized took into consideration not only the finite width of a sonar beam, but also the uncertainty associated with the vehicle pose, making evidence grids especially attractive. More recently, a vehicle equipped with single-beam sonars was used to inspect and map flooded sinkholes [4]. The 2D SLAM technique used in that application relies on particle filters and a specialized occupancy grid implementation that is amenable to the high number of operations associated with particle filter techniques.

Due to the nature of the measurements taken by these single- and multi-beam sonars, the direct use of feature-based techniques (i.e. on individual measurements) is limited. Similar to side-scan sonars, multi-beam sonars return echo intensities along rays cast from the sensor, in a plane normal to the vehicle's trajectory (i.e. a grayscale image in polar coordinates), resulting in no overlap between subsequent measurements. This, together with the apparent lack of "feature-richness", severely limits the use of "dense" techniques such as the ones previously mentioned. Submap techniques address this limitation by accumulating measurements in a higher level map representation over some limited time interval.

Small ROVs equipped with mechanically swept sonars

---

This work was partially supported by ONR grants N00014-12-1-0093 and N00014-14-1-0373.

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

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

[1] We use the term *dense* to refer to (smoothing) techniques where a new pose is instantiated with every new measurement.

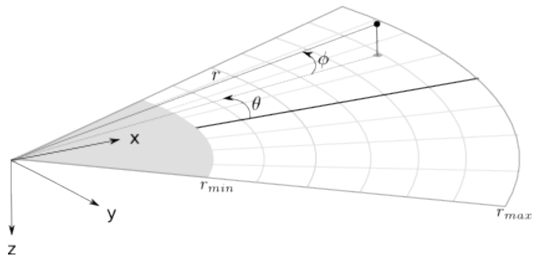


Fig. 1. Multibeam sonar: beam geometry (adapted from [2]).

have also been used to map man-made structures. One approach has the ROV hover in place while the sonar assembles a full 360-degree, two-dimensional scan of its surroundings, from which line features are extracted and utilized as landmarks in a filtering (EKF) context [5]. Alternatively, scan-matching techniques can also be used to obtain relative pose measurements [6]. This is of particular interest for applications in less structured environments, such as the inspection of ancient water cisterns [7]. In this last example, a complete 2D map is built from pairwise registration of scans. To mitigate the motion-induced errors associated with hovering, the ROV lands at the bottom of the cistern before sweeping the sonar to assemble a scan.

Submap-based methods have also been successfully applied to the problem of bathymetric mapping [8]. Registered multi-beam returns are assembled into submaps, which are checked for “richness” and consistency/drift. Like the previous examples, pairwise submap registrations—obtained through use of 2D correlation and Iterative Closest Point (ICP) algorithms—are incorporated in an EKF as relative pose measurements. A similar submap technique has also been proposed in the context of ship hull inspection, replacing filtering with smoothing by integrating pairwise registrations as loop-closure constraints in a factor graph [9]. Finally, note that other submap techniques that deal with “dense” approaches are also available in the literature—see, for instance, Tectonic SAM [10], where submaps are created by partitioning a pre-existing factor graph, and the Atlas framework [11], which aims at reducing computational effort by limiting the complexity of individual submaps.

The aim of this work is to develop a mapping technique able to produce a map of a complex underwater structure with centimeter-level resolution. This resolution requirement stems from the features of interest in the ship hull inspection scenario, which have characteristic dimensions on the order of a few centimeters to a few tens of centimeters. Additionally, it is desirable that such a framework can be run in real time, either due to the time-critical nature of some of ship hull inspection tasks, or so that the map can be used by an ROV operator for navigation. Finally, our focus lies on map-building using only sonar data for sensing, for the reasons mentioned above.

Our proposed approach builds upon previous work on submap-based mapping by combining the submap partitioning techniques of [8] with a pose graph framework similar to that of [9]. Moreover, we replace the direct registration of sonar returns—common to both approaches—with the use of volumetric techniques, preceded by enhancement and classification of sonar data. This increases robustness to outliers and classification errors, reducing the need for smoothing or filtering of submaps and improving the overall map quality.

## II. PROBLEM STATEMENT

The problem at hand, as previously mentioned, is that of inspecting underwater structures with complex geometries (e.g. curved surfaces, non-convexities, sharp edges, etc.). To accomplish this task, we assume that a hovering AUV/ROV is used. These vehicles are typically equipped with the “standard” navigation payload, comprising a Doppler Velocity Log (DVL), an Attitude and Heading Reference System (AHRS), and a depth sensor. By utilizing the attitude estimate from the AHRS, the DVL’s measurement of bottom-relative velocity in body frame coordinates can be projected onto—and integrated in—the world frame. When combined with the AHRS output, this constitutes a full 6DoF pose estimate; nevertheless, actual implementations will often replace the $z$ estimate obtained through integration with the direct measurements available from high accuracy depth sensors. Furthermore, the use of tactical grade inertial measurement units in the AHRS results in roll and pitch accuracies of better than $0.1^\circ$, reducing the pose estimation problem from 6 to 3 degrees of freedom ($x$, $y$, and $\psi$). However, unlike the AHRS roll and pitch estimates, the yaw estimate is not drift-free. While this can be mitigated through the use of an aiding sensor such as a compass, this is not always desirable, particularly when working in the proximity of metal structures such as a ship hull. As a result, the horizontal pose estimates provided by these systems are bound to drift over time. As indicated in the previous section, inspection platforms often carry a multibeam sonar as part of its primary payload. These sonars comprise an array of transducers that produce one or more narrow beams that sweep within a plane to create a scan (figure 1). The number of beams and their horizontal (i.e. “in-plane”) beam width determine the sonar’s horizontal field of view[2] and angular resolution. Range resolution is determined by the timing characteristics of the receiving circuit (longer time bins will yield lower radial resolution and vice versa). Typical values of a few hundred of beams, beam widths of $1^\circ$ or less, centimeter-level range resolution, and typical ranges on the order of a few tens of meters make them excellent sensors for underwater mapping, capable of providing high-resolution scans at large stand-off distances. Multi-beam sonars often output their measurements as a 2D array, where each element corresponds to the return intensity for that particular range and angle bin—essentially an image in polar coordinates (see figure 3(a) for an example of a sonar image after conversion to Cartesian coordinates).

Since our objective is to produce a map with a resolution such that objects of a certain size can be identified, we

---

[2] With the vertical field of view being equal to the vertical width of a beam.

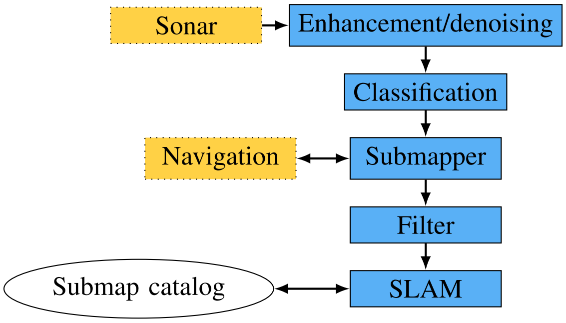


Fig. 2. Overview of the proposed mapping pipeline.

must then ensure that the uncertainty associated with the mapped sonar returns or features remains bounded and below the required resolution. This requires *(i)* robust classification/detection of valid sonar returns, *(ii)* accurate (short-term) registration of those returns, and *(iii)* long-term position accuracy through drift correction.

## III. SUBMAP CREATION

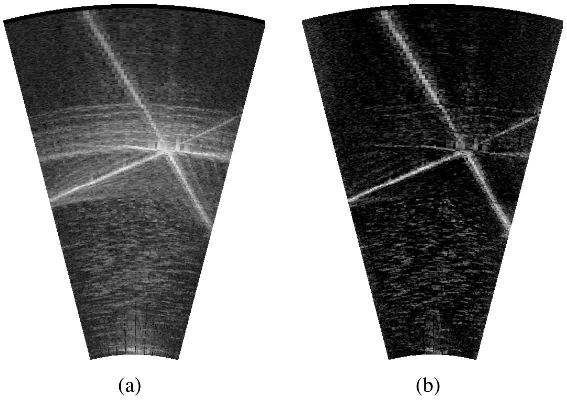
Our submap creation pipeline can be summarized as follows (figure 2): we begin by enhancing the incoming sonar data, with the aim of reducing some of the most significant systematic errors present in the data. This makes range extraction—the second step in our approach—much more reliable, seeing as there are less outliers in the data. Having extracted and registered the sonar ranges, they are inserted into an occupancy grid to build volumetric submaps, which are then converted into point clouds and filtered before being handed over to the SLAM module (section IV).

### A. Data Enhancement

Like any sensor, multibeam sonars are not free from either random or systematic error. Random error is caused both by the acoustic noise in the environment and electronic noise in the transducer and signal processing circuitry. Systematic error manifests itself in several different ways, and can be traced to a combination of causes, including non-zero beam-width, cross-talk, vehicle motion, and multi-path scattering [12]. Figure 3(a) shows a corner of a tank, as seen by the multibeam sonar used in our experiments (covered in greater detail in section V). There we observe both noise and the artifacts caused by the different error sources: angular and radial blur due to the non-zero beam-width; a mirror image of one of the walls caused by multi-path scattering, and crosstalk between transducer beams resulting in the curved ("arc"-like) feature. Considering that range extraction will rely on intensity values from the sonar scan, it is thus highly desirable to mitigate these effects as much as possible.

Our approach to this problem is to model the sonar as a linear, time-invariant (LTI) system, where the output image $y$ is the result of a convolution between the ideal image $x$ and the sonar impulse response $h$, under additive noise $n(r, \theta)$:

$$y(r, \theta) = x(r, \theta) * h(r, \theta) + n(r, \theta) \tag{1}$$



(a) (b)

Fig. 3. A sonar scan of a corner before (a) and after (b) enhancement. Sensor origin (not pictured) is located at the bottom the image, and range increases towards the top.

If this model holds, we can then try to identify the sensor impulse response (also known as *point spread function*, or PSF, in the context of imaging systems), and deconvolve the image using a two-dimensional Wiener filter [13]. To model the PSF, however, we must first understand how the sonar forms an image.

To reduce cross-talk artifacts, the DIDSON fires its 96 transducers in a staggered fashion—it takes 8 cycles, each firing 12 transducers, to form a complete image. This strategy limits cross-talk to the beams firing in the same cycle, which, despite being separated by 7 inactive transducers, are still "sensitive" to each other, as can be seen from figure 4.

If we assume the other transducers in the array have the same beam pattern, then the PSF can be obtained by sampling the beam pattern at the angular positions of the other transducers firing in the same cycle. Note that this approach has some limitations in the sense that it *(i)* only captures the *angular* component of the PSF, and, *(ii)* assumes the PSF is isotropic. While addressing the former assumption would likely require some prior information about the scene being mapped, we can at least account for the angular dependence of the PSF by pre-multiplying the sonar image by an angle-dependent function to account for the fact that the beam pattern for a transducer, while keeping the same overall shape, decreases in amplitude as we move away from the center of the array. A similar approach is used to compensate for the radial decay in return intensity, by modeling both geometric and absorption losses. The results of this enhancement step are shown in figure 3.

### B. Range extraction

Having enhanced the sonar image, the next step is to classify it, that is, to detect the presence (or absence) of objects in the current scene. We follow a standard technique [5], [9], [14], [15] for multibeam sonar data by selecting, for each beam, either the first or the strongest return that is above a certain threshold. For the $i^{th}$ detected return, at coordinates $(r_i, \theta_i)$, the 3D position of the return in the sonar frame is obtained by conversion to Cartesian coordinates: $x^s = [r_i \cos(\theta_i), r_i \sin(\theta_i), 0]^T$. Using homogeneous coordinates,

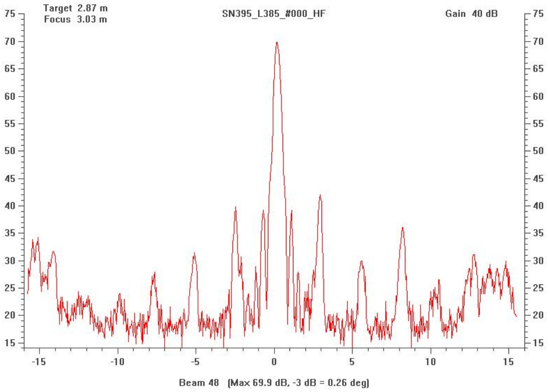


Fig. 4. Typical beam pattern for one of the DIDSON's beams—this will vary slightly over different sonars and lenses. (figure courtesy of Sound Metrics Corp., used with permission).

the location of that return in the global frame is

$$\tilde{x}^g = T_v^g T_s^v \tilde{x}^s \tag{2}$$

where $T_s^v$ is the (fixed) transform describing the sonar pose in the vehicle frame, and $T_v^g$ is the vehicle pose estimate in the global frame, obtained from the navigation solution.

### C. Volumetric mapping

Despite the improvements in return selection obtained through the enhancement of sonar data, the classifier still has non-zero probabilities for both false positives (false detections) and false negatives (missed detections). This can be further aggravated by the presence of bubbles, suspended sediments, and fish schools, which, due to the strong sonar returns they produce, will also be registered, effectively resulting in outliers. In order to mitigate the undesirable effects of such outliers, we use probabilistic volumetric techniques, where instead of registering returns, we update the occupancy probability for the region of space where that return lies. These techniques have the additional advantage that they let us use *negative information*—that is, we can use the knowledge that the space between the vehicle and the return is empty to update the occupancy probability for the regions traversed by the beam. To compensate for the non-zero beam width and height of the sonar beams, we have extended the *OctoMap* library [16] to support sonar beam insertion—this allows us to implicitly model some of the angle uncertainty associated with a sonar return and, since a larger number of voxels have their occupancy probabilities updated, obtain denser submaps.

### D. Point clouds and filtering

Once a submap has been assembled, it is transformed to a point cloud by thresholding: for each voxel whose occupancy probability exceeds a user-defined level, a point is added to the corresponding point cloud, with its coordinates being equal to the center of the voxel. Since some outliers will still make their way through the volumetric submapping module, subsequent filtering is often necessary. For this purpose we employ a clustering-based filter: points are first grouped into clusters based on each point's distance to their nearest neighbours(s), and small clusters (below a given number of points) are then removed. The motivation for this filter stems from the observation that these "persistent" outliers are, in fact, the result of creatures and/or objects (e.g. fish, bubbles) that remain in the sonar's field of view over the course of a few frames, leaving a small, sparse cloud of points in their wake.

### E. Submap span

On the one hand, submaps should be large enough so as to have enough geometric features that allow for a submap to be registered against another; on the other, a submap must be short enough that the accumulated odometry error is small. Under linear motion at constant depth, and given measurement covariances of $(\sigma_x^2, \sigma_y^2)$ for the horizontal velocity and $\sigma_\psi^2$ for the yaw angle, the pose covariance after $T_i$ seconds is

$$\Sigma_i = T_i \cdot diag(\sigma_x^2, \sigma_y^2, \sigma_\psi^2) \tag{3}$$

The two main assumptions in the equation above—linear motion and reduction to 3 dimensions—will be addressed in greater detail in the following sections.

## IV. 3D SLAM

Using range measurements extracted from enhanced sonar scans, we have assembled a sequence of submaps connected by odometry-based relative pose estimates. Knowing that these estimates are not drift-free, if we want to improve navigation performance and map accuracy we must leverage the information contained in the submaps. As we have previously mentioned, one way to do this is through *pairwise submap registration*: for any submap pair for which there is scene overlap, we can obtain a new relative pose constraint by aligning the two so that these match in the overlapping area. While this can be a better estimate of relative pose, this is not always the case, especially for submaps with small overlap and/or simple geometries (e.g. mostly planar). As we obtain more of these estimates, we arrive at an optimization problem where we must try and find the pose estimates that best fit both odometry and pairwise registration constraints—we tackle this problem using a *factor graph* approach.

### A. Factor graph

At every instant $t_i$, a new pose estimate $x_i$ and sonar scan $S_i$ are available[3]. The relative pose between two subsequent scans can be written as:

$$u_{i,i+1} = x_{i+1} \ominus x_i \tag{4}$$

This can be represented as a factor graph, where each pose estimate $x_i$ is associated with a node, and odometry constraints $u_{i,i+1}$ are represented as edges connecting the nodes for poses $x_i$ and $x_{i+1}$. The assoicated covariance, $\Sigma_i$, can be determined using equation (3). At this point,

---

[3] Even though the sonar is likely not to output a new scan at the same time or frequency as the navigation solution is computed, the latter can be interpolated to obtain a pose estimate at the time a scan was received.

the (acyclic) factor graph represents what is known as an "odometry chain" (figure 5), and the value of each pose can be obtained through composition of odometry constraints.

### B. Submapping in a factor graph perspective

In this factor graph representation, each node is implicitly paired with the measurements taken at that instant—in our case, the sonar scans. For the reasons listed in section I, these are rarely informative enough that loop closure information can be extracted from a pair of scans; instead, subsequent scans are "accumulated" into a finite set (represented by the colors in figure 5(a)), so that constraints can then be obtained by pairwise analysis of these sets. This "accumulation" process—the assembly of volumetric submaps from individual sonar scans, introduced in section III-C—can be described as the registration of all sonar scans in one *submap*, in the frame of the first pose of that submap: the base pose (figure 5(b)). The edges connecting the base poses of submaps $j$ and $j+1$ are obtained by composition of the odometry constraints of the poses associated with submap $j$.

### C. Pairwise registration

At this point we have arrived at a higher-level odometry chain, where each base node is associated with a submap instead of a sonar scan (figure 5(c)); still, if navigation performance is to be improved, additional constraints need to be added to the factor graph. While detection and matching of 3D features (such as planes) from submap pairs may be a feasible technique, we intend to keep our technique capable of dealing with both structured and unstructured environments—thus, we use scan matching to derive these loop closure constraints.

Once a new submap is assembled, we begin by identifying potential matches by finding other submaps with which it overlaps. For each potentially matching submap a transformation estimate is computed through ICP [17], [18]. Since all points in each submap are referenced to its base pose, we compose the odometry constraints connecting the two nodes to arrive at the initial transformation estimate. Having obtained a transformation estimate $o_{ij}$ for the alignment between submaps $S_i$ and $S_j$, we use the resulting score, as well as the time difference between submap acquisitions, to provide us with some guidance on its acceptance (e.g. the greater the time difference, the greater the threshold on the difference between odometry and ICP transformations, as more drift is likely to have been accumulated). Given a process model $x_i = f(x_{i-1}, u_{i-1})$ subject to additive white Gaussian noise (AWGN) $w_i$ with covariance, and a measurement model $o_{jk} = h(x_j, x_k)$, also under AWGN $v_{jk}$, we want to find the *maximum a posteriori* (MAP) estimate $X = \{x_1, \ldots, x_N\}$:

$$X_{MAP} = \arg\min_{X} \left( \sum_{i=1}^{N} \|f(x_{i-1}, u_{i-1}) - x_i\|^2 + \sum_{(j,k)\in O} \|h(x_j, x_k) - o_{jk}\|^2 \right) \tag{5}$$

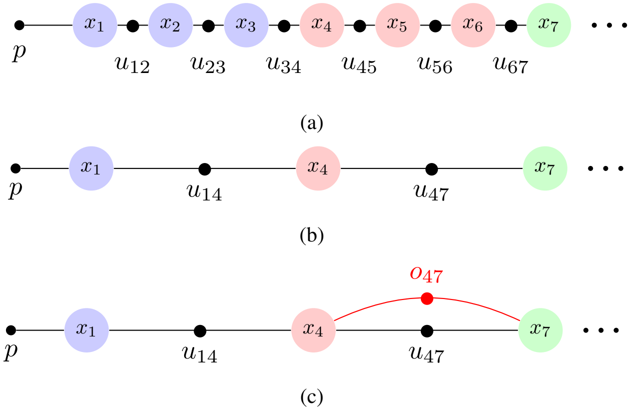


Fig. 5. An overview of our submapping technique from a factor graph perspective: the starting graph (a) for the robot trajectory has each pose $x_i$ connected to the next pose $x_{i+1}$ by an edge $u_{i,i+1}$ (odometry constraint) with associated covariance $\Sigma_i$; poses are grouped into submaps, with the corresponding *base poses* replacing the originating poses (b); finally, pairwise constraints are added between base poses whose corresponding submaps have been co-registered (c). Note that the covariance matrices associated with each edge are not represented.

Fig. 6. The Hovering Autonomous Underwater Vehicle.

where $\Lambda_i$ is the covariance matrix for the process model noise, and $O$ is the set of all tuples $(j,k)$ for which a pairwise registration constraint $o_{jk}$ has been added between $x_j$ and $x_k$, with associated covariance matrix $\Gamma_{jk}$. We use iSAM [19] to solve this optimization problem and obtain updated base pose estimates $x_i$. The final map is constructed by merging all submaps once their base poses have been replaced with the updated values.

## V. EXPERIMENTAL RESULTS

In this section we begin by briefly describing the inspection vehicle and its navigation and sensor payloads before going over the results obtained for the two datasets.

### A. Platform

In our experiments we use the *Hovering Autonomous Underwater Vehicle* (HAUV) [20] from MIT/Bluefin (figure 6). It is equipped with five rim-driven thrusters that make it

TABLE I
NAVIGATION PAYLOAD: SENSOR PERFORMANCE

| Sensor | Axis | Accuracy | Unit |
|---|---|---|---|
| IMU | Roll, Pitch | 0.02 | ° |
|  | Yaw | 0.05 | ° |
| DVL | x,y,z | 0.003 | m/s |
| Depth | z | 0.01 | % |

directly controllable in all axes but roll. Its navigation payload comprises a Honeywell HG1700 inertial measurement unit (IMU), a 1.2MHz Teledyne/RDI Workhorse Navigator DVL, and a Paroscientific Digiquartz depth sensor—the relevant performance characteristics are summarized in table I. The DVL can rotate parallel to the vehicle's pitch axis, allowing for both bottom- and hull-relative motion.

The HAUV's primary payload is a dual-frequency identification sonar (DIDSON) [21], which we use primarily in its high frequency (1.8MHz) mode, as it provides greater detail. It has a total of 96 beams, each with 0.3° of horizontal beam width. Since we are interested in complex geometries, we make use of a *concentrator* lens, reducing its vertical field of view from the standard 14° to 1° (-3dB values). While its range resolution depends on the minimum and maximum range configuration, a typical window of 9.5 meters will yield a resolution of better than 2 centimeters.

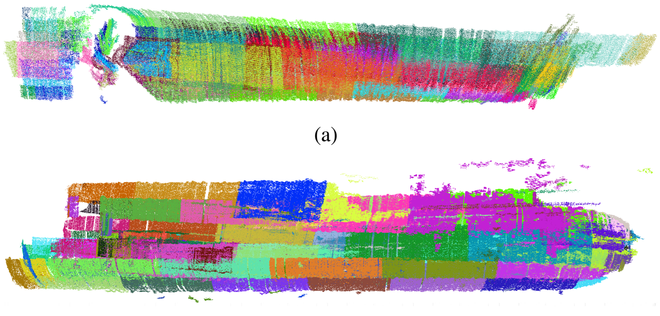
### B. SS Curtiss dataset

This dataset covers the inspection of the *complex areas* of the *SS Curtiss*: the running gear (rudder and propeller) and the bulbous bow (highlighted areas in figure 7). Both datasets were acquired while performing "coarse" surveys—essentially a vertical lawn-mower pattern (*i.e.* constant depth transects) designed to rapidly create a map for the purposes of situational awareness. This is of particular use on unseen ships or structures, before any close-in inspection is performed. Since resolution is not as important as coverage and coverage rate, the vehicle is deployed with the sonar scan plane parallel to the hull's longitudinal cross-section plane. Submap duration was chosen to be equal to 30 seconds, which, for the performance specifications listed in table I, corresponds to a horizontal uncertainty of approximately 0.02m and 3° ($1\sigma$ values). Finally, the results shown were obtained in real time while performing inspection.

Fig. 7. Inspected regions of the hull of the *SS Curtiss* for the *running gear* (red) and *bulbous bow* (blue) datasets.

*1) Running gear:* This dataset comprises a "coarse" survey of the running gear from the starboard side, covering the rudder, propeller, and the hull sections immediately forward of the latter. The coverage and results for this run are shown in figures 8(a) and 10, respectively.

By performing horizontal transects parallel to the ship's longitudinal axis, we are able to quickly (approximately 30 minutes) cover one side of the running gear. However, the relatively fast coverage rate is not without disadvantages: the complex geometry of the propeller is poorly represented as the propeller is not imaged from its front or back, resulting in sparse coverage of each blade. Unlike the hull and rudder, which are nearly normal to the incident sonar beams, propeller blades are ensonified at shallow angles, resulting in increased radial uncertainty in the sonar returns.



(a)

(b)

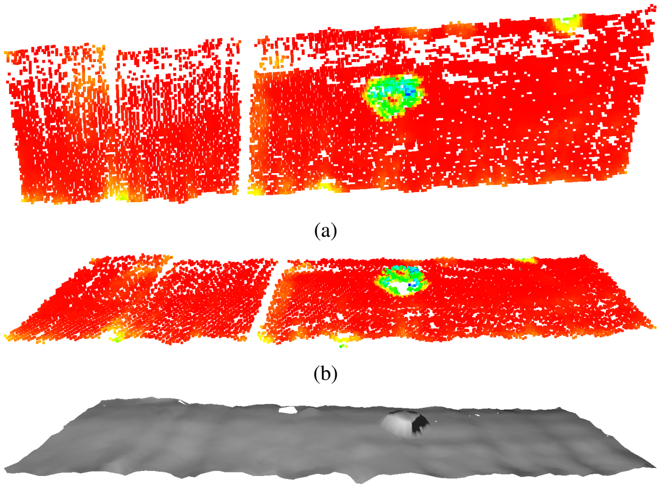
Fig. 8. Submap coverage for the running gear (a) and bulbous bow (b) datasets (each colored patch represents a submap). The maps span approximately 45 and 39 meters, respectively. Map height is roughly equal to the ship's draft: 9 meters.

Despite the relatively short duration of this inspection run, there is still a significant amount of drift in the vehicle's horizontal position estimates. This is especially evident on the forward and aft edges of the map: the rudder appears as multiple parallel sections (instead of a single, continuous surface), with the distance between these sections reaching 2 meters. The proposed SLAM technique is able to reduce this worst-case distance to approximately 70cm while improving the overall shape of the map. This improvement is also noticeable towards the forward section of the map where this inter-submap distance is reduced from 1.2m to 40cm (approximate values).

*2) Bulbous bow:* A "coarse" survey of the bulbous bow of the ship, again from starboard side, was also performed and captured in this dataset. The coverage and results for this run are shown in figures 8(b) and 11, respectively.

Like the previous dataset, inspection of the bow was done by performing horizontal transects parallel to the ship's longitudinal axis. Due to the curvature of the hull, however, and in order to keep a consistent stand-off distance from the hull, a small yaw maneuver was performed about halfway through each transect. This resulted in a rather large amount of drift, as can be seen from the forward and aft regions of the map. On the top half of figure 11, it is possible to distinguish three "instances" of the bow tip—this has been successfully corrected by the proposed SLAM technique (figure 11, bottom).

This inspection run also includes detailed inspection of a test object placed on the hull prior to the inspection run, and found in the course of the survey. This results obtained from a sample submap containing this object are shown in figure 9. From these we can see that while the shape has not been perfectly recovered (the object is a round metal



(a)

(b)

(c)

Fig. 9. Submap with test object from the bulbous bow dataset: curvature-coded point clouds (profile (a) and perspective (b) views) and resulting mesh (c). Submap dimensions are approximately 1.3 by 5 meters; object dimensions (as measured from the submap) are 11x32x22 centimeters.

cylinder 30cm in diameter), the submap is accurate enough to represent features on the scale of a few centimeters. Also of note is the vertical gap in figures 9(a) and 9(b)—this is the result of a missed sonar frame, possibly due to network congestion.

## VI. CONCLUSION AND FUTURE WORK

We have proposed a technique for sonar-based SLAM that, while tailored to underwater scenes with complex geometries, can also be used in other applications such as bathymetric mapping. This technique extends previous work by focusing on improving submap accuracy through modeling of systematic errors and the use of probabilistic volumetric techniques for increased outlier robustness. Experimental validation on the problem of ship hull inspection indicates that it is both able to improve map accuracy by successfully correcting for navigation drift and represent features on the order of a few centimeters.

In our approach to sonar data enhancement (section III), we limited modeling to the angular component of the point spread function and made the assumption that it was angle-invariant. Relaxing these limitations should, in principle, yield better results: adding a radial dimension would improve range accuracy. Furthermore, the experimentally observed angular variation in beam pattern could be accounted for in an anisoplanatic PSF. Despite the relative success in eliminating outliers and improving overall submap quality, the use of volumetric techniques (as well as the subsequent filtering step) points to the challenge of sonar image classification. Our classifier is biased towards “empty space” (instead of occupied) to reduce the number of outliers at the cost of missing actual returns—because of its simple threshold-based model, increasing its *sensitivity* would necessarily increase the number of outliers. Thus, improving classification performance is one of the main goals of ongoing work. Still related to the classification problem, and motivated by the end-use of the sonar in a mapping framework, is the possibility of extracting additional information from sonar data, namely, surface orientation: given the narrow vertical field of view of an imaging sonar, it should be possible to at least restrict (if not completely determine) the orientation of surface elements to a finite set. This would be of particular use to the registration (ICP) step in our SLAM approach, and to the reconstruction of a 3D mesh from the resulting set of maps—the data product of an inspection system.

Additional improvements to the accuracy of both the SLAM solution and final map could be obtained by delaying the conversion from occupancy grids to point clouds to further along in the mapping pipeline. Replacing the use of ICP in the pairwise registration step with dense volumetric alignment techniques would improve pose estimate accuracy as both positive and negative information are used in the process. The final map could then be obtained by merging the different SLAM-corrected volumetric submaps onto a single occupancy grid and only then reconstructing a point cloud and/or 3D mesh. Finally, another relevant topic for future work would be the extension of these techniques to the inspection of dynamic scenes.

## REFERENCES

[1] P. Ozog and R. M. Eustice, “Real-time SLAM with piecewise-planar surface models and sparse 3D point clouds,” in *Int. Conf. Intelligent Robots and Syst. (IROS)*. IEEE, 2013, pp. 1042–1049.

[2] H. Johannsson, M. Kaess, B. Englot, F. Hover, and J. Leonard, “Imaging sonar-aided navigation for autonomous underwater harbor surveillance,” in *Int. Conf. Intelligent Robots and Syst. (IROS)*. IEEE, 2010, pp. 4396–4403.

[3] W. Stewart, “A non-deterministic approach to 3-D modeling underwater,” in *Proc. 5th Int. Symp. Unmanned Untethered Submersible Technology*, vol. 5. IEEE, 1987, pp. 283–309.

[4] N. Fairfield, G. Kantor, D. Jonak, and D. Wettergreen, “Autonomous exploration and mapping of flooded sinkholes,” *Int. J. of Robotics Research*, vol. 29, no. 6, pp. 748–774, 2010.

[5] D. Ribas, P. Ridao, J. Neira, and J. D. Tardos, “SLAM using an imaging sonar for partially structured underwater environments,” in *Int. Conf. Intelligent Robots and Syst. (IROS)*. IEEE, 2006, pp. 5040–5045.

[6] A. Mallios, P. Ridao, D. Ribas, and E. Hernandez, “Probabilistic sonar scan matching SLAM for underwater environment,” in *Oceans*. IEEE, 2010, pp. 1–8.

[7] 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,” in *Int. Conf. Robotics and Biomimetics (ROBIO)*. IEEE, 2012, pp. 538–544.

[8] C. Roman and H. Singh, “A Self-Consistent Bathymetric Mapping Algorithm,” *J. of Field Robotics*, vol. 24, no. 1-2, pp. 23–50, 2007.

[9] M. VanMiddlesworth, M. Kaess, F. Hover, and J. Leonard, “Mapping 3D Underwater Environments with Smoothed Submaps,” Brisbane, Australia, Dec. 2013.

[10] K. Ni, D. Steedly, and F. Dellaert, “Tectonic SAM: Exact, Out-of-Core, Submap-based SLAM,” in *Int. Conf. Robotics and Automation (ICRA)*. IEEE, 2007, pp. 1678–1685.

[11] M. Bosse, P. Newman, J. Leonard, and S. Teller, “Simultaneous localization and map building in large-scale cyclic environments using the Atlas framework,” *Int. J. of Robotics Research*, vol. 23, no. 12, pp. 1113–1139, 2004.

[12] M. VanMiddlesworth, “Toward autonomous underwater mapping in partially structured 3D environments,” Master’s thesis, Massachusetts Institute of Technology, Feb. 2014.

[13] E. E. Hundt and E. A. Trautenberg, “Digital processing of ultrasonic data by deconvolution,” *Trans. on Sonics and Ultrasonics*, vol. 27, no. 5, pp. 249–252, 1980.

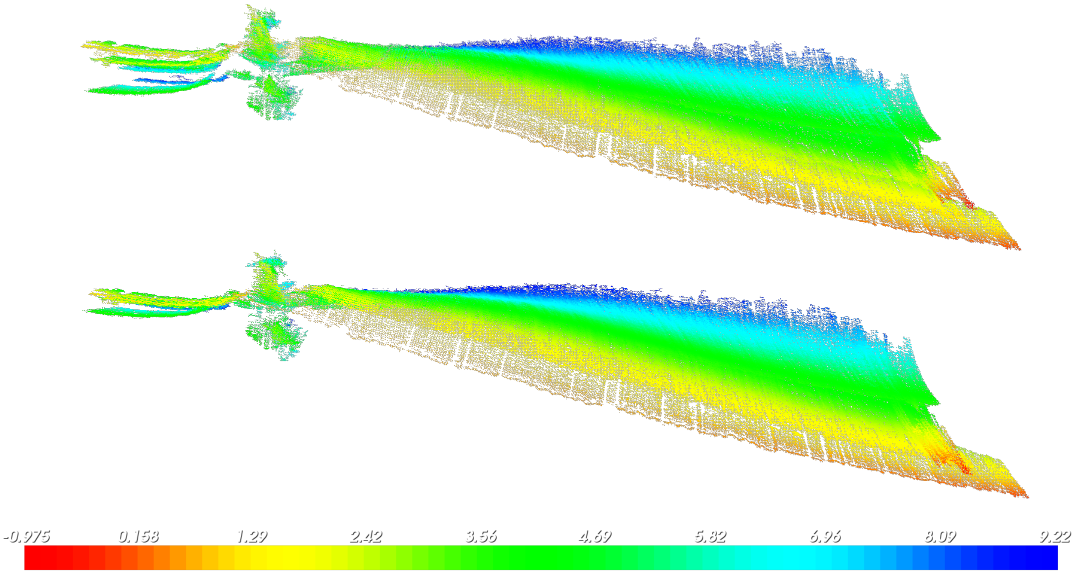


Fig. 10. Top view of the map of the running gear with depth-based color coding: odometry (top) and SLAM (bottom). This map was obtained through constant depth transects, with the vehicle running parallel to the ship's longitudinal axis, and spans approximately 45 meters.

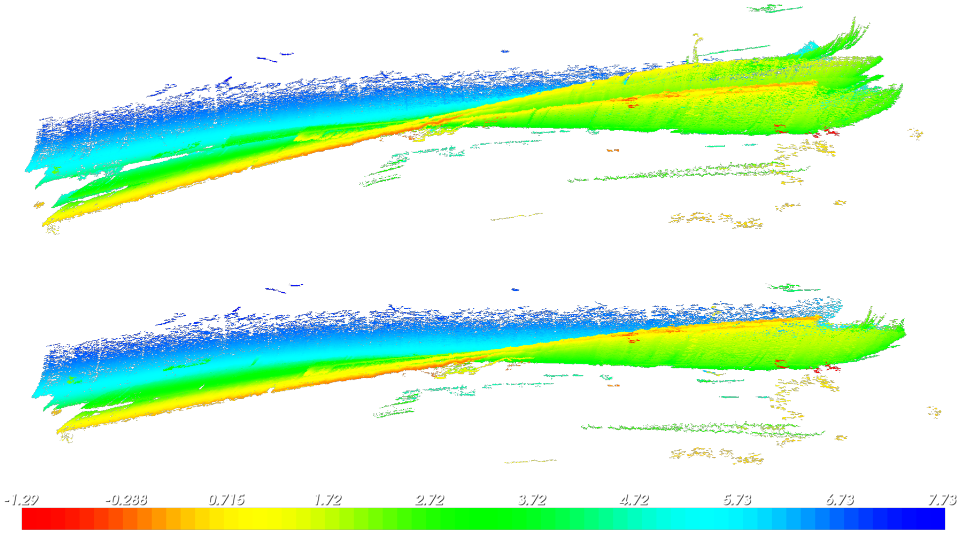


Fig. 11. Top view of the map of the bulbous bow with depth-based color coding: odometry (top) and SLAM (bottom). This map was obtained through constant depth transects, with the vehicle running parallel to the ship's longitudinal axis and performing a yaw maneuver approximately halfway through the transect. The map spans approximately 39 meters.

[14] C. Roman and H. Singh, "Improved vehicle based multibeam bathymetry using sub-maps and SLAM," in *Int. Conf. Intelligent Robots and Syst. (IROS)*. IEEE, 2005, pp. 3662–3669.

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

[16] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, "OctoMap: An efficient probabilistic 3D mapping framework based on octrees," *Autonomous Robots*, 2013. [Online]. Available: http://octomap.github.com

[17] P. J. Besl and N. D. McKay, "Method for Registration of 3-D Shapes," in *Robotics-DL tentative*. Int. Society for Optics and Photonics, 1992, pp. 586–606.

[18] R. B. Rusu and S. Cousins, "3D is here: Point Cloud Pibrary (PCL)," in *Int. Conf. Robotics and Automation (ICRA)*. IEEE, 2011, pp. 1–4.

[19] M. Kaess, A. Ranganathan, and F. Dellaert, "iSAM: Incremental Smoothing and Mapping," *Trans. on Robotics*, vol. 24, no. 6, pp. 1365–1378, 2008.

[20] 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 J.*, vol. 41, no. 2, pp. 44–55, 2007.

[21] E. Belcher, W. Hanot, and J. Burch, "Dual-frequency identification sonar (DIDSON)," in *Proc. Int. Symp. Underwater Technology*. IEEE, 2002, pp. 187–192.