Pose-graph SLAM using Forward-looking Sonar

Jie Li, Michael Kaess, Ryan M. Eustice and Matthew Johnson-Roberson

***Abstract*—This paper reports on a real-time simultaneous localization and mapping (SLAM) algorithm for an underwater robot using an imaging forward-looking sonar (FLS) and its application in the area of autonomous underwater ship hull inspection. The proposed algorithm overcomes specific challenges associated with deliverable underwater acoustic SLAM, including feature sparsity and false-positive data association when utilizing sonar imagery. Advanced machine learning technique is used to provide saliency aware loop closure proposals. A more reliable data association approach using different available constraints is also developed. Our evaluation is presented on real-world data collected in a ship hull inspection application, which illustrates the system's performance and robustness.**

***Index Terms*—Marine robotics, Localization, SLAM**

I. Introduction

RECENT research development has seen dramatically increasing performance in autonomous underwater vehicles (AUVs). Platforms have improved in efficiency, computational capacity and sensor quality. AUVs have the potential to revolutionize our access to the oceans, addressing a variety of critical problems such as climate change assessment [1], marine habitat monitoring [2], and underwater structure inspection [3]–[5].

However, navigation and localization of AUVs is still a challenging problem mainly due to the strong attenuation of electromagnetic signals, including the global positioning system (GPS), in the aqueous environment. One of the most significant problems faced by these platforms is the unconstrained drift that occurs without any external positional observations. Theoretically, the unbounded drifting can be corrected by two types of approaches: calibration through external infrastructure (surveyed acoustic beacons) or self-correction using environmental landmarks, which is part of the well studied simultaneous localization and mapping (SLAM) algorithm employed in many current systems [6]. While the mathematical theory of the SLAM problem has been extensively explored and reliable optimization approaches have been proposed to solve it, open practical problems still exist. In the underwater domain, due to the aforementioned electromagnetic attenuation, both *gathering* and *associating* sensor measurements are challenging. Due to the limited propagation of light and signal sensitivity to water clarity as well as illumination conditions, more common and well-studied sensors in terrestrial SLAM systems such as optical cameras and LIDAR do not work well in some of the underwater environments. In contrast, acoustics have a long and rich history as perceptual sensors in the underwater domain [7]–[9], as it offers greater range and insensitivity to water visibility. A two-dimensional acoustic sensor, forward-looking sonar (FLS), is a promising alternative to optical sensing in underwater SLAM systems against turbid water environment. Nevertheless, the utilization of this modality suffers from unique challenges including low image texture and high signal-to-noise ratio (SNR).

In this paper, we propose a novel pose-graph SLAM algorithm leveraging FLS as the sole perceptual sensor to provide ego-motion drift correction. In this work, we address many of the practical problems associated with leveraging signals from FLSs, including feature sparsity, low reliability in data association, and singularity in geometry estimation. Specifically, we propose the following novel contributions: 1) Creation of a system to identify and select the most informative sonar frames for use in improving system efficiency and reliability, while avoiding the singularities which pervade the geometry estimation problem when using FLS; 2) development of a robust sonar feature matching strategy using multiple constraints including pose prior within a SLAM framework; 3) evaluation of the efficacy of our algorithm for the application scenario of underwater structure inspection and assessment. This evaluation is done on several sets of large-scale real-world data. This data was collected in a ship hull inspection project [3], [10], [11]. Experimental results show that the proposed algorithm is able to provide robust loop closure detection and relative pose constraint estimation from FLS, which ultimately minimizes drift in vehicle localization.

The rest of the paper is arranged as follows: In Section II, we give a brief introduction on related works using FLS in AUV navigation; in Section III, we give an overview of the ship hull inspection project by which the proposed algorithm is motivated; in Section IV, we provide a thorough introduction to the proposed algorithm; in Section V, we evaluate the proposed algorithm through the use of ship hull inspection data; and finally, in Section VI, we conclude the work discussing promising future directions.

II. Background

A majority of research work in FLS imaging focuses on image registration between two sonar image frames. [12]

Manuscript received: September 10, 2017; Revised December 15, 2017; Accepted February 7, 2018.

This paper was recommended for publication by Editor Jonathan Roberts upon evaluation of the Associate Editor and Reviewers' comments. This work was supported in part by the Office of Naval Research under award N00014-16-1-2102 & N00014-16-1-2103, and by the Department of Energy under award DE-EM0004383.

J. Li was with the Department of Electrical Engineering & Computer Science, University of Michigan, Ann Arbor, MI, USA. She is now with Toyota Research Institute, Ann Arbor, MI, USA jie.li@tri.global.

Michael Kaess is with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA kaess@cmu.edu.

R. Eustice and M. Johnson-Roberson are with the Department of Naval Architecture & Marine Engineering, University of Michigan, Ann Arbor, MI, USA {eustice,mattjr}@umich.edu.

Digital Object Identifier (DOI): see top of this page.

provides a thorough review on recent registration methods of sonar images. Popular approaches span from spectral methods [13] to feature-based methods [7]. These approaches have shown promising results in pairwise image registration and motion estimation under certain assumptions on scene geometry and sufficient frame overlapping. However, these methods are mainly targeted at the application of image mosaicing. It is non-trivial to adapt them to real-time navigation systems due to computational complexity.

Pioneering research has been conducted to enable the use of FLS in AUV navigation systems. Walter et al. [14] proposed the first SLAM implementation using the Exactly Sparse Extended Information Filter (ESEIF) with manually selected features from sonar imagery. Building the foundation of FLS for SLAM, this work focused more on the efficiency in building a feature map to ensure real-time performance and less on feature detection or addressing geometric ambiguities. Johannsson, Kaess, Englot, *et al.* [15] proposed detecting and matching FLS image features using Normal Distribution Transformation (NDT). This work makes a planar assumption where all feature points are assumed to lie on a plane parallel to the vehicle. This assumption helps to add extra constraints to address the ambiguity in elevation measurements. However, it also limits the application scenarios in which such an algorithm can be used.

A similar assumption was made by Shin, Lee, Choi, *et al.* in their recent work on FLS-based bundle adjustment for seabed mapping [16]. This work introduces the Accelerated-KAZE (A-KAZE) [17] feature for acoustic sonar images and displays impressive performance on the challenging data association problem. Fallon, Folkesson, McClelland, *et al.* [18] also proposed an FLS-based SLAM system for the seabed surveying scenario. They proposed to use multi-hypothesis analysis to reacquire features extracted through seabed subtraction. Since the data association is performed based on feature locations, the approach is more adapted for feature sparse environments.

In more recent literature, Huang and Kaess [8] introduced Acoustic Structure From Motion (ASFM), which used multiple sonar viewpoints to jointly reconstruct 3D structures as well as sensor motion, with no specific assumption of the scene. In their work, statistical analysis was conducted to identify and discuss the degeneration cases for which ASFM would fail. Particularly, they analyze which motion patterns make for a poorly constrained optimization problem. In another recent work on acoustic-inertial odometry estimation, Yang and Huang [9] extended the discussion of this problem to more abstract theoretical cases. Although ASFM shows promising results for providing sonar-based geometry estimation without underlying scene assumptions, this work was primarily evaluated on simulated data.

However, despite ongoing work there are still open issues, particularly when it comes to the practical deployment of an acoustic SLAM system using FLS. These problems include, but are not limited to, robust data association, efficient key frame selection for real-time performance, and identification and removal of degeneration cases in the optimization. This paper is complementary of these works on ASFM. We present

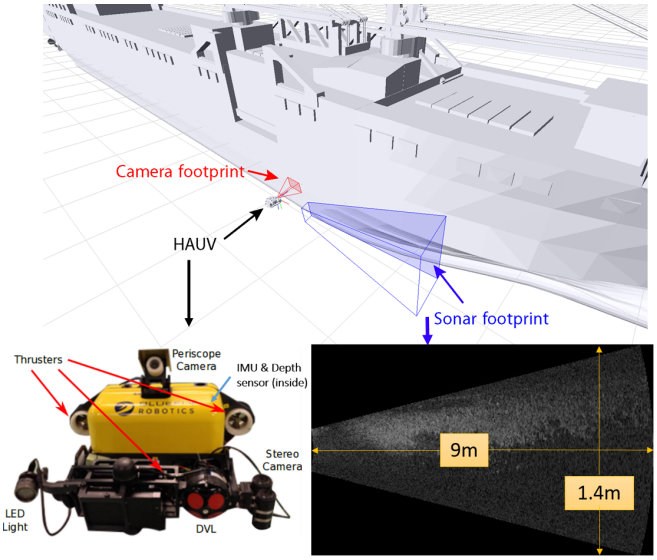


Fig. 1: HAUV sensor payload and its optical/acoustic footprint with respect to a ship hull during the inspection.

novel algorithm that makes use of ASFM as front-end and develop the missing pieces for a complete end-to-end real-time acoustic SLAM system. The hurdles to creating a deployable framework that can run on an AUV are addressed, providing solutions to the practical problems discussed above.

III. Ship Hull Inspection using HAUV

We concretize the proposed SLAM system for use in a long-term periodic ship hull inspection project using a Hovering Autonomous Underwater Vehicle (HAUV) [3], [19]. In this project, an HAUV from Bluefin Robotics [20] is used (Fig. 1). The vehicle is equipped with a hull-looking 1200 kHz RDI Doppler velocity log (DVL) and Honeywell HG1700 IMU providing odometry measurements. For perception, the vehicle is equipped with a set of stereo cameras for underwater vision and a periscope camera for above water vision. The vehicle is also equipped with a 1.8MHz DIDSON Sonar [21] to provide acoustic-based perception looking to the right of the vehicle as depicted in Fig. 1. During the inspection, the HAUV follows a lawn mower pattern with respect to the ship hull keeping a fixed standoff of $1m$.

Previous work has been done within the project to establish a real-time navigation system using visual sensors [3], [19]. This is one of the key functionalities for the HAUV to perform the inspection in a location-aware manner, which also provides the basic requisition for active path planning in different scenarios [22]. This work is mainly motivated by improving the real-time navigation robustness and reliability in camera inactive conditions, such as turbid water environment.

IV. Methodology

In this section, we describe the proposed acoustic SLAM system in detail. An overview of the approach is given in Fig. 2. A pose-graph SLAM back-end is used to estimate the vehicle's full 6 degree of freedom (DOF) poses, $\{\mathbf{x}\}_i = \{[x, y, z, \phi, \theta, \psi]\}_i$ given all constraints between poses estimated from sensor measurements. Fig. 3 depicts

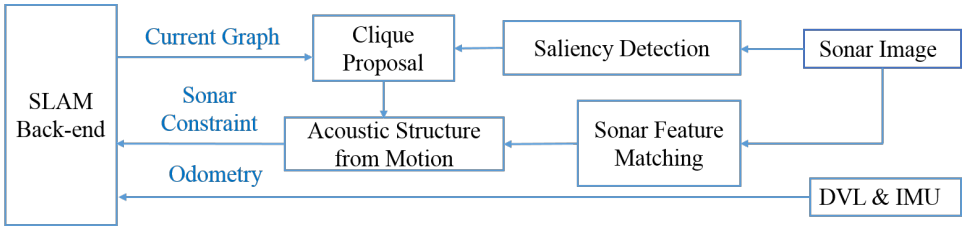


Fig. 2: System flowchart: The proposed algorithm is supported by a pose-graph SLAM back-end. It optimizes the vehicle poses given all constraints sensor measurements. Raw sonar frames are passed through a saliency test where salient frames are considered for loop closure proposal based on potential information gain. Marginalized pose constraints, after local ASFM, will be fed to the SLAM back-end providing drift correction from sonar measurements.

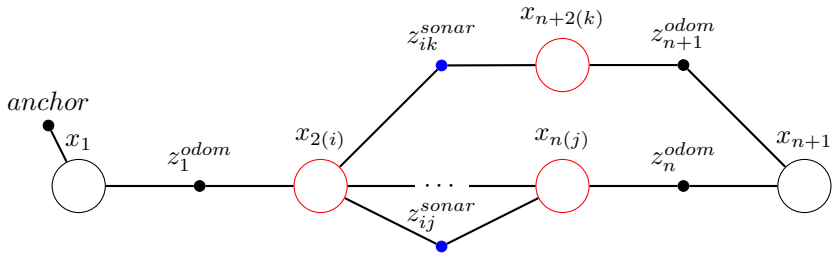


Fig. 3: Graphical model of pose-graph SLAM using constraints from local FLS bundle adjustment. An example of loop closure clique is marked in red (i, j, k). The constraint factors from local ASFM are colored in blue.

the general graphical model of our pose-graph representation. Each node in the graph, $\mathbf{x}_i$, corresponds to a vehicle pose with a sensor measurement associated. Two types of constraints from sensor measurements are considered in this context: odometry measurements from the DVL and inertial measurement unit (IMU), and sonar constraints from the FLS local ASFM (described later in detail in Section IV-B). We assume independent Gaussian noise for both odometry and sonar measurements. The measurement model for odometry is given by:

$$\mathbf{z}_i^{odom} = f(\mathbf{x}_i, \mathbf{x}_{i+1}) + \mathbf{w}_i, \tag{1}$$

where $f()$ denotes the generative model of odometry between two poses and $\mathbf{w}_i$ is Gaussian noise with its covariance scaled proportional to the elapsed time separating the poses $\mathbf{x}_i$ and $\mathbf{x}_{i+1}$.

To model the constraints provided by FLS images, we propose to use marginalized pose constraints from a local ASFM problem within a clique of loop closure poses. We will give a brief introduction of sonar geometry and sonar-based ASFM in Section IV-A and Section IV-B. Then we give more details about how we provide an advanced loop closure hypothesis proposal and feature matching that improves the robustness of the sonar based ASFM in Section IV-C and Section IV-D respectively.

A. Sonar Frame Geometry

The FLS sends out acoustic waves across the field of view (FOV). Then an array of transducers observes reflected signal from the scene and returns measurements in the form of range and bearing (r, θ). The limitation of this type of sonar lies in the ambiguity in elevation (ϕ). The measurement for each data point (r, θ) can originate anywhere on the arc defined by (r, θ) in the spherical coordinate system centered at the projection center of the sonar head, as depicted in Fig. 4.

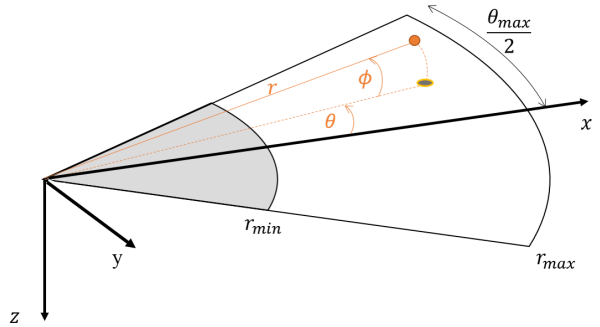


Fig. 4: Forward-looking sonar geometry.

In a local Cartesian coordinate system as defined in Fig. 4, the coordinates of a 3D scene point $\mathbf{P}$ can be converted from the spherical coordinates as follows:

$$\mathbf{P} = \begin{bmatrix} X_s \\ Y_s \\ Z_s \end{bmatrix} = \begin{bmatrix} r\cos\theta\cos\phi \\ r\sin\theta\cos\phi \\ r\sin\phi \end{bmatrix} \tag{2}$$

The raw measurement of $\mathbf{P}$, (r, θ), can be converted to a 2D image for easier interpretation and post processing. The specific manner in which image conversion occurs can vary depending on the application scenario. In this work, we leverage one of the most common: the Cartesian coordinate image conversion. Under this conversion, a 3D point $\mathbf{P}$ is projected on the image plane $(\phi = 0)$:

$$\mathbf{p} = \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \gamma \cdot Y_s/cos\phi + \frac{w}{2} \\ \gamma \cdot (r_{max} - X_s/cos\phi) \end{bmatrix} \tag{3}$$

where (u, v) is the pixel location centered at the upper left corner of the sonar image and $[w, h]$ are the width and height of the image. γ is a scaling factor between the image pixel space and the physical space, which is given by:

$$\gamma = \frac{w}{2r_{max}\sin(\theta_{max}/2)}, \tag{4}$$

where θ_{max} is the sonar FOV of bearing.

B. FLS Structure from Motion

Given the geometry of imaging sonars described in Section IV-A, we can build a local ASFM problem within a small set of sonar frames by modeling both the odometry measurements and sonar feature measurements. We consider a three-frame ASFM in the implementation. It is straight forward to generalize the work over any number of frames.

The ASFM solves for the maximum *a posteriori* (MAP) vehicle poses and landmark positions $\Theta = \{\mathbf{x}_i, \mathbf{P}_m\}$ given all the measurements $\mathbf{Z} = \{\mathbf{z}_{ij}^{prior}, \mathbf{z}_{im}^{sonar}\}$ in a given clique:

$$\Theta^* = \text{argmax}_\theta \, p(\Theta|\mathbf{Z}) = \text{argmax}_\theta \, p(\Theta)p(\mathbf{Z}|\Theta). \quad (5)$$

In the ASFM setting, the poses in a clique are not necessarily sequential. Instead of direct odometry measurements from the sensors, we use relative pose-to-pose constraint from the current SLAM estimate, given by:

$$\mathbf{z}_{ij}^{prior} = f(\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_j) + \mathbf{w}_{ij}^{prior}, \quad (6)$$

where $\bar{\mathbf{x}}_i$ and $\bar{\mathbf{x}}_j$ are pose estimates from the SLAM back-end, and $\Sigma_{\mathbf{w}_{ij}^{prior}}$ is given by the propagated covariance of $f(\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_j)$. We define sonar measurement as:

$$\mathbf{z}_{im}^{sonar} = h(\mathbf{P}_m, \mathbf{x}_i) + \mathbf{v}_{im}, \quad (7)$$

where $h()$ denotes the generative model that converts a landmark position $\mathbf{P}_m$ to the local frame of vehicle pose $\mathbf{x}_i$, then projects the landmark on the image plane following Eq. 3. $\mathbf{v}_{im}$ denotes the Gaussian noise associated with the feature measurement on the sonar frame. We model the measurement uncertainty proportional to the radius of the detected key point. The whole ASFM is then a solution to the least squares problem of:

$$\Theta^* = \text{argmin}_\Theta [\Sigma_{m=1}^{M} \Sigma_{i=1}^{N} ||h(\mathbf{P}_m, \mathbf{x}_i) - \mathbf{z}_{im}^{sonar}||_\Sigma^2 + \Sigma_{i=1}^{N} \Sigma_{j=i+1}^{N} ||f(\mathbf{x}_i, \mathbf{x}_j) - \mathbf{z}_{ij}^{prior}||_\Sigma^2], \quad (8)$$

where M is the number of landmarks that can be detected in all the sonar frames in the ASFM clique and N is the number of poses in the clique. Mahalanobis distance ($||\mathbf{x}||_\Sigma^2 = \mathbf{x}^T \Sigma^{-1} \mathbf{x}$) is used to measure all the residuals. Note one implementation detail is that we inflate the uncertainty of the relative odometry constraint in the z-axis to compensate for the spatial ambiguities in the biased motion pattern particular to the ship hull dataset.

After the optimization, we take the marginalized odometry constraints as new pose constraints for the SLAM back-end, which is given by:

$$\mathbf{z}_{ij}^{ASFM} = f(\mathbf{x^*}_i, \mathbf{x^*}_j) + \mathbf{w}_{ij}^{ASFM}. \quad (9)$$

C. Saliency Aware Loop Closure Proposal

To ensure real-time performance of the system, an efficient hypothesis proposal system is essential. This system will propose loop closures of sonar frame cliques to build a local ASFM problem. The selection of these loop closure cliques is based upon the evaluation of both potential information gain given successful registration and the estimated saliency of the underlying sonar frames (a proxy for potential matching quality).

The use of information gain for link proposal was introduced by Ila, Porta, and Andrade-Cetto [23] in which one only adds informative links to keep the SLAM graph compact. Kim and Eustice [3] also extend this approach for loop closure hypothesis proposal between camera images. The information gain of a new measurement between two nodes under the assumption of a jointly Gaussian distribution is given by:

$$L = \mathbf{H}(\mathbf{X}) - \mathbf{H}(\mathbf{X}|\mathbf{z}_{ij}) = \frac{1}{2} \ln \frac{|\mathbf{S}|}{|\mathbf{R}|}, \quad (10)$$

where $\mathbf{H}(\mathbf{X})$ and $\mathbf{H}(\mathbf{X}|\mathbf{z}_{ij})$ is the prior and posterior entropy of the graph. $\mathbf{R}$ is the measurement covariance and $\mathbf{S}$ is the innovation covariance.

$$\mathbf{S} = \mathbf{R} + [\mathbf{H}_i, \mathbf{H}_j] \begin{bmatrix} \Sigma_{ii} & \Sigma_{ij} \\ \Sigma_{ji} & \Sigma_{jj} \end{bmatrix} [\mathbf{H}_i, \mathbf{H}_j]^T \quad (11)$$

The above equation can be extended to the case of several measurements:

$$\mathbf{S}_{i,j,k}^{12\times12} = \mathbf{R}_{i,j,k}^{12\times12} + \mathbf{H} \begin{bmatrix} \Sigma_{ii} & \Sigma_{ij} & \Sigma_{ik} \\ \Sigma_{ji} & \Sigma_{jj} & \Sigma_{jk} \\ \Sigma_{ki} & \Sigma_{kj} & \Sigma_{kk} \end{bmatrix}^{18\times18} \mathbf{H}^T \quad (12)$$

where $\mathbf{R}$ here becomes the covariance of the new measurement $\mathbf{z}_{new} = (\mathbf{z}_{ij}, \mathbf{z}_{ik})$ and $\mathbf{H}$ is the corresponding Jacobian matrix of $\mathbf{z}_{new}$. They are given by:

$$\mathbf{R}_{i,j,k} = \begin{bmatrix} \Sigma_{\mathbf{z}_{ij}\mathbf{z}_{ij}} & \Sigma_{\mathbf{z}_{ij}\mathbf{z}_{ik}} \\ \Sigma_{\mathbf{z}_{ik}\mathbf{z}_{ij}} & \Sigma_{\mathbf{z}_{ik}\mathbf{z}_{ik}} \end{bmatrix} \quad (13)$$

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_i^{\mathbf{z}_{ij}} & \mathbf{H}_j^{\mathbf{z}_{ij}} & \mathbf{0} \\ \mathbf{H}_i^{\mathbf{z}_{ik}} & \mathbf{0} & \mathbf{H}_k^{\mathbf{z}_{ik}} \end{bmatrix} \quad (14)$$

Eq. 10 provides an evaluation of potential information gain that will contribute the SLAM optimization problem. Another advantage of the information gain metric is that it encourages the cliques with higher complexity in motion pattern, which decreases the chance of a singularity in the local ASFM problem, as discussed in [8], [9].

However, as we discussed in Section I, the sparse distribution of sonar features is a fundamental challenge in making a successful ASFM constraint. Proposals based *solely* on information gain can lead to a lot of failed attempts, due to lack of features, decreasing the efficiency of the entire system. To address this problem, we combine both information gain and the sonar image saliency to determine which images to use for proposing loop closures.

We analyze sonar frame saliency based upon the location sensitivity of a global image feature learned using a Convolutional Neural Network (CNN). In the authors' previous work [24], we proposed a global image feature descriptor using a CNN that provides informative features that helps vehicle localization. Fig. 5 depicts the data flow for training and run-time use of the feature model. For more details of the model specifics the reader is referred to [24]. In this work, we propose an on-line saliency evaluation based on local sensitivity of the CNN trained image feature given by:

$$Sa_j = \frac{\sqrt{\Sigma_{i=1}^{N_j} ||\mathbf{f}_i - \bar{\mathbf{f}}||_2^2}}{\sqrt{\Sigma_{i=1}^{N_j} ||\mathbf{p}_i - \bar{\mathbf{p}}||_2^2}}, \ \mathbf{p}_i \in B_j, \quad (15)$$

where Sa_j is the saliency score of frame j and B_j is the support area of frame j. B_j includes all the N_j frames that contain overlapping observation areas with frame j. $\mathbf{p}_i = (x, y, z)_i$ is the vehicle location and $\bar{\mathbf{p}}$ denotes the mean vehicle location in B_j. $\mathbf{f}_i$ is a 3000 dimensional feature

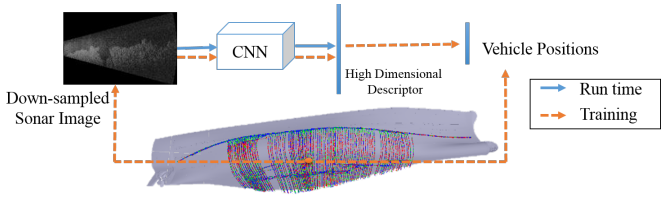


Fig. 5: Image-level descriptive sonar feature training procedure proposed in [24].

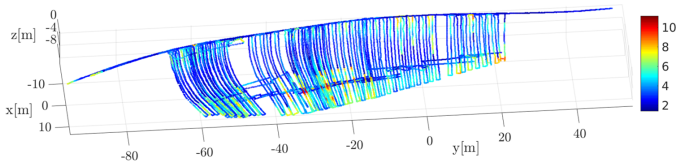


Fig. 6: Saliency score estimated from image-level sonar feature with respect to the vehicle position.

descriptor extracted from frame i and $\bar{\mathbf{f}}$ denotes the mean descriptor in B_j. The saliency score captures the variability of sonar features $\mathbf{f}_i$ within a fixed physical support area around frame j. Fig. 6 depicts the distribution of the scores evaluated online as the vehicle moves with respect to the environment (in this case the ship hull). Examples of the sonar frames with associated saliency scores are given in Fig. 7. As can be seen, the saliency score presents correlation with sonar image texture diversity.

We use the proposed saliency score to prune the loop closure proposals by evaluating information gain within only the salient frames:

$$L^s(i,j,k) = \begin{cases} L(i,j,k) = \frac{1}{2}\ln\frac{|S_{i,j,k}|}{|R_{i,j,k}|} & \overline{Sa} > \lambda_s, \\ 0 & otherwise. \end{cases} \quad (16)$$

$\overline{Sa}$ denotes the mean saliency score in the clique (i,j,k) and λ_s is a cut-off threshold for the saliency score. $\lambda_s = 5$ was experimentally found to be optimal for the data gathered in the ship hull dataset. The saliency aware information gain $L^s(i,j,k)$ is used to propose and prioritize the cliques considered in the local ASFM front-end.

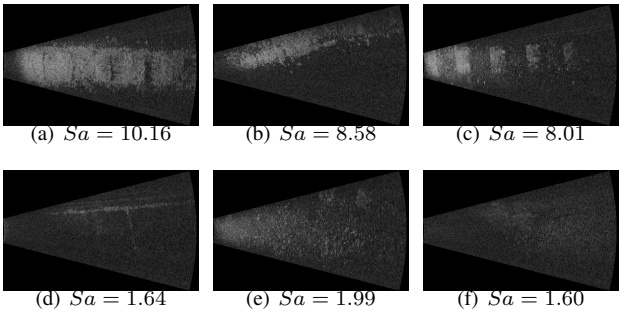
Since the proposal of a clique of size m is a $O(n^m)$ problem, we use a sampling-based method for the initial proposals and feed them into a fixed-size queue prioritized by $L^s(i,j,k)$.

In Section V, we show that the saliency aware information gain helps to prioritize the proposals that could construct a well constrained local ASFM problem.

D. Sonar feature matching using constraint

The problem of constructing feature point associations between sonar frames is one of the most challenging tasks in using an imaging sonar in a pose estimation and reconstruction problem. Inspired by the work in [25] that leveraged pose constraints in underwater image feature matching, we propose to constrain the search area of feature matching by modeling the uncertainty and constraints of all the variables in the feature generative model. For a feature point $\mathbf{p}_m^i$ detected on frame i, we expect the corresponding feature point location in frame j to be defined as:

$$\mathbf{p}_m^j = h(\mathbf{P}_m(\mathbf{p}_m^i, \phi_m), \mathbf{x}_i, \mathbf{x}_j), \quad (17)$$



(a) $Sa = 10.16$ (b) $Sa = 8.58$ (c) $Sa = 8.01$

(d) $Sa = 1.64$ (e) $Sa = 1.99$ (f) $Sa = 1.60$

Fig. 7: Example sonar frames with saliency scores. (a) (b) and (c) give examples of high saliency sonar frames, where we can see strong texture and features. (d) (e) and (f) give examples of low saliency frames, where little texture or features can be recognized.

where ϕ_m is the elevation of underlying a 3D scene point $\mathbf{P}_m$ and $\mathbf{x}_i, \mathbf{x}_j$ are the corresponding vehicle poses. Given the elevation ϕ_m and feature point $\mathbf{p}_m^i$, $\mathbf{P}_m$ can be uniquely reconstructed following Eq. 2 and Eq. 3.

From Eq. 17, we can approximate a Gaussian probability distribution of $\mathbf{p}_m^j$ using linear covariance propagation:

$$\mu_{\mathbf{p}_m^j} = h(\mathbf{P}_m(\mathbf{p}_m^i, \phi_m), \mathbf{x}_i, \mathbf{x}_j)|_{\phi_m=0} \quad (18)$$

$$\Sigma_{\mathbf{p}_m^j} = \mathbf{J}\Sigma_{\mathbf{p}_m^i,\phi_m,\mathbf{x}_i,\mathbf{x}_j}\mathbf{J}^T, \quad (19)$$

Where $\mathbf{J}$ is the Jacobian matrix of function h. $\Sigma_{\mathbf{p}_m^i,\phi_m,\mathbf{x}_{ij}}$ denotes the joint covariance matrix of $\mathbf{p}_m^i, \phi_m, \mathbf{x}_{ij}$. Note that $\mathbf{p}_m^i, \phi_m, (\mathbf{x}_i, \mathbf{x}_j)$ are independent variables. We can simplify Eq. 19 by:

$$\Sigma_{\mathbf{p}_m^j} = \mathbf{J}_{\mathbf{p}_m^i}\Sigma_{\mathbf{p}_m^i}\mathbf{J}_{\mathbf{p}_m^i}^T + \sigma_{\phi_m}^2\mathbf{J}_{\phi_m}\mathbf{J}_{\phi_m}^T + \mathbf{J}_{\mathbf{x}_i\mathbf{x}_j}\Sigma_{\mathbf{x}_i\mathbf{x}_j}\mathbf{J}_{\mathbf{x}_i\mathbf{x}_j}^T. \quad (20)$$

$\sigma_{\phi_m}^2 = \phi_{max}$ is the covariance of ϕ_m since we have no cue of ϕ_m due to the ambiguity. $\Sigma_{\mathbf{x}_i\mathbf{x}_j}$ is the joint covariance of pose frame $\mathbf{x}_i$ and frame $\mathbf{x}_j$ available from SLAM. The radius of the feature scale of the keypoint is used to give the covariance of the detected feature point $\Sigma_{\mathbf{p}_m^i}$.

$$(\mathbf{p}_m^j - \mu_{\mathbf{p}_m^j})^T\Sigma_{\mathbf{p}_m^j}^{-1}(\mathbf{p}_m^j - \mu_{\mathbf{p}_m^j}) = \kappa^2, \quad (21)$$

where κ^2 follows a χ_2^2 distribution.

A set of example matching results is given in Fig. 8 and it is also compared to nearest neighborhood matching. A-KAZE detection and speeded up robust features (SURF) feature descriptors are used in the implementation. It can be seen that our matching approach results in many more correct associations. While some outliers still exist, in the improved matching approach, the ratio of inliers is high enough to be effective in an ASFM framework. In the implementation of the system an initial ASFM is conducted to enable the identification of gross outliers that can be excluded from a second ASFM pass. Only the feature points that can be detected in every frame in the clique are used in the ASFM.

V. Experiment

In this section, we evaluate our proposed algorithm on real-world data collected in a ship hull inspection application. We provide quantitative and qualitative results evaluating the vehicle localization accuracy of the proposed SLAM algorithm using FLS as the only perceptual sensor. During the ship hull inspection, the HAUV is operating in a GPS-denied environment with no external reference set up. No ground

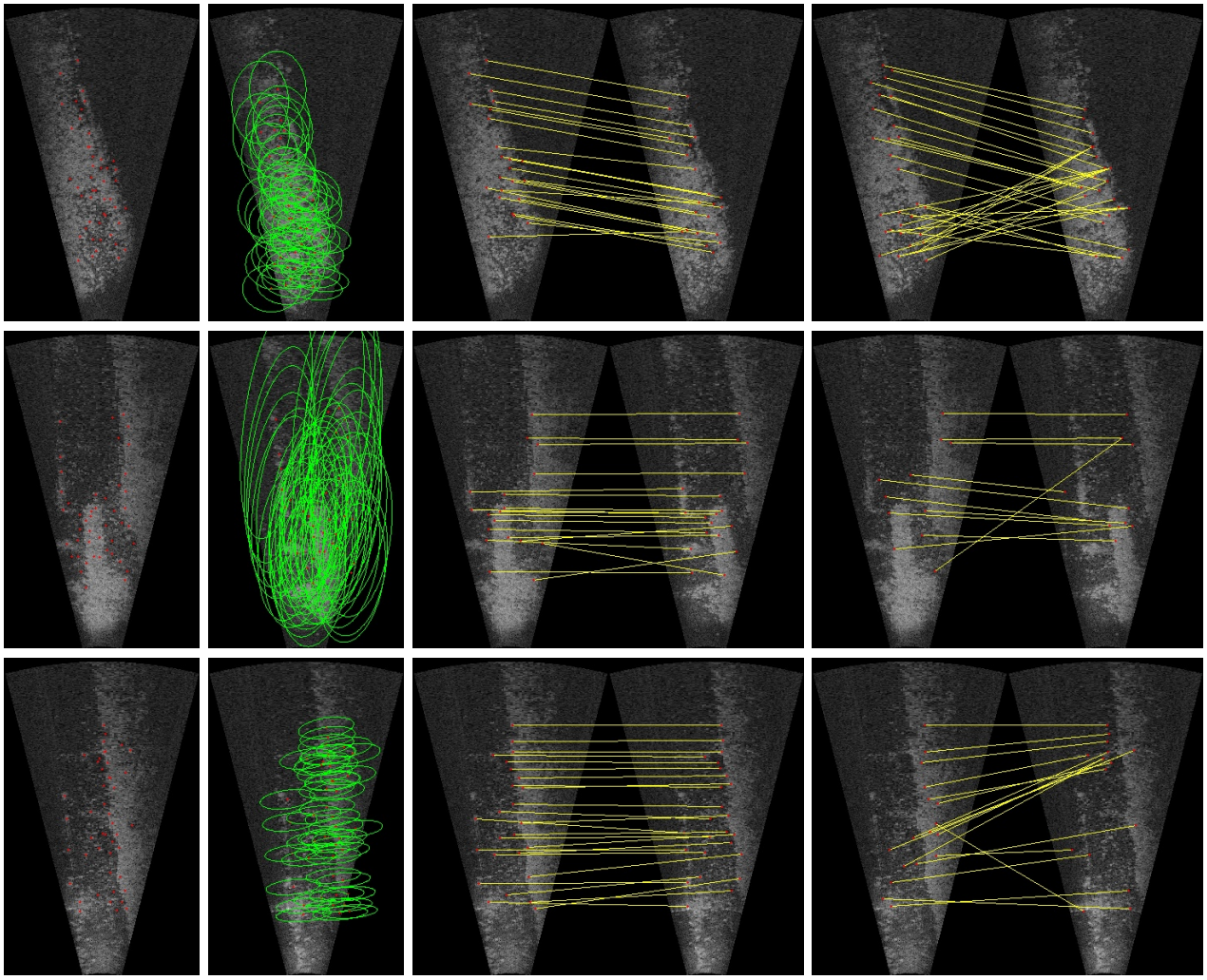


Fig. 8: Sonar feature matching using constraints. The first column gives sonar image i with detected feature points marked in red. The second column gives sonar frame j. Expected correspondent location of the feature points in frame j is marked in red on frame j. Searching area in frame j is marked in green. The third column gives feature matching results using a constrained searching area. The fourth column gives feature matching results using the nearest neighboring matching searching in the whole image.

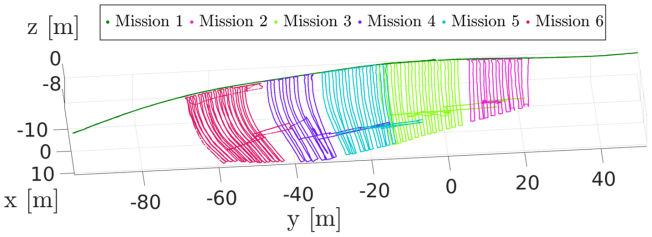


Fig. 9: SLAM missions collected in the inspection of the SS-Curtiss in 2014. This figure depicts six missions in a common reference frame.

TABLE I: SLAM inference statistics. We record the process time with or without Saliency information (Sa.) in the loop.

Mission		1	2	3	4	5	6
Number of frames		2240	2210	4389	2987	2597	5440
Number of salient frames		113	609	1142	569	874	1037
Mission duration [min]		40.0	36.3	74.6	49.0	46.6	98.8
Process time [min]	w/o Sa.	72.1	76.3	170.0	108.1	93.9	252.2
	w/ Sa.	**4.3**	**22.11**	**43.0**	**22.3**	**33.3**	**47.1**
Speed up over real time	w/o Sa.	0.55x	0.48x	0.43x	0.45x	0.49x	0.39x
	w/ Sa.	**9.30x**	**1.64x**	**1.73x**	**2.20x**	**1.40x**	**2.09x**

truth is available on vehicle localization. To provide better evaluation of the proposed method, we compare the estimated trajectory with recovered vehicle poses using an off-line CAD model-assisted bundle adjustment framework reported by Ozog and Eustice in [10]. This offline method leverages visual measurements as well as other on-board sensors with exhaustive feature correspondence searching, which provides a reasonable benchmark on the localization accuracy for this application under clear water quality. We also compare the results to dead reckoning localization as a baseline.

The dataset used in our experiment comes from the inspection of the SS-Curtiss in 2014. A set of trajectories around different portions of the ship hull is included in the experiment, focused on areas that exclude the rudder and screws. Fig. 9 depicts the spatial distribution of each trajectory. Detailed statistics of each trajectory as well as SLAM inference information are summarized in Table I.

All the experiments are conducted on a Lenovo laptop with an Intel(R) Core(TM) i7-3840QM CPU@2.80GHz and a Quadro K2000M GPU. The sonar image descriptive feature is extracted on a separate thread through the GPU, while everything else is implemented single threaded on the CPU. We use Ceres Solver [26] as the optimization solver. The

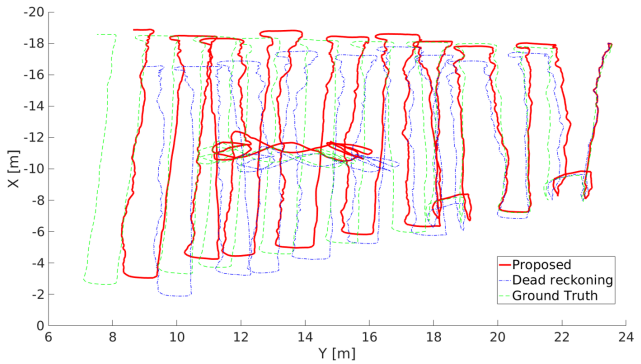


Fig. 10: Trajectory estimation result of Mission 2. As the drift accumulates, the constraints from the sonar local structure from motion are able to correct the trajectory towards the benchmark trajectory. Our propose method provides promising drifting correction on the X axis in this single mission. On the Y axis, correction is presented when the error is relatively large, but it is less obvious than the correction on the X axis.

process efficiency meets the real time requirement based on the current implementation, as shown in Table I.

Table II depicts SLAM localization accuracy with respect to benchmark positions compared with dead-reckoning trajectories. To account for the variability introduced by the randomization of the sample-based initial proposal selection method introduced in Section IV-C, we conduct 20 trial runs for each mission. The mean and standard deviation of the absolute error in each trial are reported.

It can be told from Table II that the incorporation of the sonar constraint is able to improve the localization accuracy in most of the missions. The improvement becomes more obvious when the relative error is higher, mainly due to the limited resolution of the sonar image with respect to the corresponding physical sensor footprint. To better illustrate how the constraint from FLS is helping to correct the drift, the SLAM trajectories of **Mission 2** compared to other methods are given in Fig. 10 with its sonar constraints depicted in Fig. 11. It can be seen that the benefit of the FLS local structure from motion front end becomes more obvious as the accumulated error increases. It is also observed from both Fig. 10 and the overall statistics results that the proposed algorithm presents better performance on the X axis comparing to the Y axis. This is mainly due to the different spatial resolution captured in the sonar image frames for our specific application. As depicted in Fig. 1, much lower spatial resolution is captured in the image on the Y axis as the HAUV performs the inspection with consistent distance and heading towards the ship hull.

a few failure cases exist where the sonar constraint is not able to improve the localization accuracy. For example, in **Mission 1**, the vehicle is performing a surface mission at the water line of the ship hull. In this type of mission, motion patterns within any loop closure cliques are close to pure translation, which is a singular pattern that causes failure to ASFM as discussed in [8]. This is an example of how, given the properties of acoustic imaging, sonar constraints are only useful for recovering gross relative motion.

It should be noted that the corrections applied by the sonar are more modest than typical SLAM corrections using optical image approaches. This is mainly due to the much lower

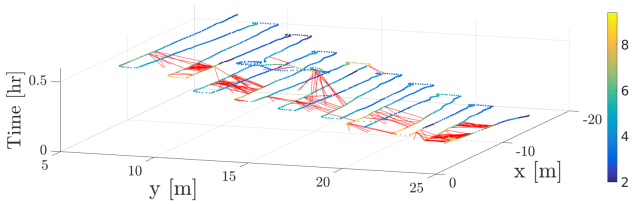


Fig. 11: FLS structure from motion constraints added to SLAM graph. The results are shown in a time elevation map. Nodes in the trajectory are colored by the saliency score associated with the sonar frames.

spatial resolution of the sonar, in the application of ship hull inspection. This finding highlights that if water clarity permits, optical constraints should be leveraged with a high priority. Ultimately, we envision that the system should be able to make use of both constraints within different parts of the same mission depending on the environment, relative hull position, and imaging geometry.

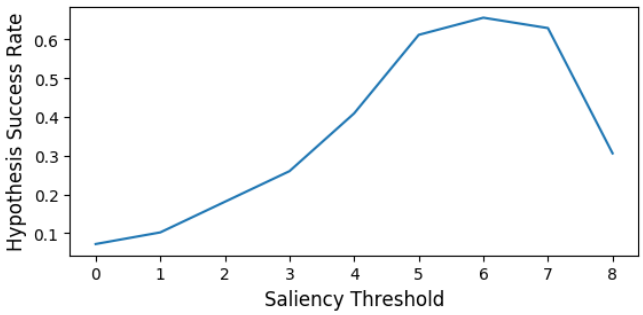


Fig. 12: Success rate of hypothesis proposals with respect to saliency threshold.

We also provide analysis on the loop closure clique proposal approach by the success rate of all hypothesis proposals when different saliency thresholds are used. Given a proposed hypothesis clique, feature correspondence searching is conducted and a local ASFM problem is created based on the matched features. If the optimization of the resulting ASFM problem converges, we mark the proposal as a successful proposal as it provides sufficient feature association and motion diversity to support a well-constrained ASFM problem. The success rate of hypothesis proposals with respect to the saliency threshold is given in Fig. 12. The results show that the saliency aware hypothesis proposal method is able to effectively suppress proposals in the low saliency area, which provides sufficient system efficiency to support real-time navigation as also depicted in Table I. If the threshold of the saliency score is too high, however, the system becomes over-selected and the ASFM success rate will drop resulting from lack of motion variety among the frames.

VI. Conclusion

In this paper, we proposed a real-time acoustic SLAM system for underwater robot navigation, which makes use of the measurements from FLS images. We have demonstrated a complete end-to-end acoustic SLAM system, particularly addressing the practical challenges of using FLS features, including data association, key frame selection and outlier rejection. The proposed method is evaluated on a real-world

TABLE II: SLAM estimation accuracy with respect to camera-image-based offline bundle adjustment for each mission. In this table we compare the localization accuracy of our proposed algorithm (Prop.) and the dead reckoning (DR). For the proposed method, the means (μ) and standard deviations (σ) of absolute error over 20 runs for each mission are given ($\mu[\sigma]$). The smallest error for each mission in each axis is in bold.

Mission		1	2	3	4	5	6
Error in X [m]	**Prop.**	**0.65[0.16]**	**0.39[0.13]**	**1.29[0.36]**	**0.39[0.15]**	**0.62[0.30]**	**2.19[0.25]**
	DR	0.79	0.52	1.64	0.68	1.06	2.44
Error in Y [m]	**Prop.**	**0.66[0.02]**	**0.74[0.11]**	0.57[0.10]	**1.49[0.27]**	**0.61[0.14]**	**1.26[0.31]**
	DR	0.70	0.90	**0.52**	1.64	0.62	1.70
Error in Z [m]	**Prop.**	1.30[0.24]	**0.42[0.07]**	0.97[0.09]	0.42[0.09]	**0.50[0.08]**	**2.08[0.23]**
	DR	**0.99**	0.50	**0.88**	**0.29**	0.57	2.57

dataset collected for a ship hull inspection application. Experimental results indicate that our approach is able to efficiently propose information rich cliques of sonar frames and make use of them in a real-time SLAM system to correct the drift in robot's position estimation. To improve the current method, an explicit singular motion pattern detection mechanism and an active path planning algorithm to encourage information rich constraints are worthy to be considered in future work. New types of point-based features specifically designed or trained for FLS images is an interesting direction to further explore this problem.

REFERENCES

[1] M. A. Moline, S. M. Blackwell, C. Von Alt, B. Allen, T. Austin, J. Case, N. Forrester, R. Goldsborough, M. Purcell, and R. Stokey, "Remote environmental monitoring units: An autonomous vehicle for characterizing coastal environments," *J. Atmospheric and Oceanic Tech.*, vol. 22, no. 11, pp. 1797–1808, 2005.

[2] S. B. Williams, O. R. Pizarro, M. V. Jakuba, C. R. Johnson, N. S. Barrett, R. C. Babcock, G. A. Kendrick, P. D. Steinberg, A. J. Heyward, P. J. Doherty, I. Mahon, M. Johnson-Roberson, D. Steinberg, and A. Friedman, "Monitoring of benthic reference sites: Using an autonomous underwater vehicle," *IEEE Robot. Autom. Mag.*, vol. 19, no. 1, pp. 73–84, 2012.

[3] A. Kim and R. M. Eustice, "Real-time visual SLAM for autonomous underwater hull inspection using visual saliency," *IEEE Trans. Robot.*, vol. 29, no. 3, pp. 719–733, 2013.

[4] J. Li, R. M. Eustice, and M. Johnson-Roberson, "Underwater robot visual place recognition in the presence of dramatic appearance change," in *Proc. IEEE/MTS OCEANS Conf. Exhib.*, 2015, pp. 1–6.

[5] ——, "High-level visual features for underwater place recognition," in *Proc. IEEE Int. Conf. Robot. and Automation*, 2015, pp. 3652–3659.

[6] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: Part I," *IEEE Robot. Autom. Mag.*, vol. 13, no. 2, pp. 99–110, 2006.

[7] M. Aykin and S. Negahdaripour, "On feature extraction and region matching for forward scan sonar imaging," in *Proc. IEEE/MTS OCEANS Conf. Exhib.*, 2012, pp. 1–9.

[8] T. A. Huang and M. Kaess, "Towards acoustic structure from motion for imaging sonar," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots and Syst.*, 2015, pp. 758–765.

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

[10] P. Ozog and R. M. Eustice, "Large-scale model-assisted bundle adjustment using Gaussian max-mixtures," in *Proc. IEEE Int. Conf. Robot. and Automation*, 2016, pp. 5576–5581.

[11] P. Ozog, M. Johnson-Roberson, and R. M. Eustice, "Mapping underwater ship hulls using a model-assisted bundle adjustment framework," *Robot. and Auton. Syst.*, vol. 87, pp. 329–347, 2017.

[12] N. Hurtos, S. Nagappa, X. Cufi, Y. Petillot, and J. Salvi, "Evaluation of registration methods on two-dimensional forward-looking sonar imagery," in *Proc. IEEE/MTS OCEANS Conf. Exhib.*, 2013, pp. 1–8.

[13] N. Hurtos, X. Cuf', Y. Petillot, and J. Salvi, "Fourier-based registrations for two-dimensional forward-looking sonar image mosaicing," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots and Syst.*, 2012, pp. 5298–5305.

[14] M. Walter, F. Hover, and J. Leonard, "SLAM for ship hull inspection using exactly sparse extended information filters," in *Proc. IEEE Int. Conf. Robot. and Automation*, 2008, pp. 1463–1470.

[15] H. Johannsson, M. Kaess, B. Englot, F. Hover, and J. J. Leonard, "Imaging sonar-aided navigation for autonomous underwater harbor surveillance," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots and Syst.*, Taipei, Taiwan, Oct. 2010, pp. 4396–4403.

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

[17] P. F. Alcantarilla and T Solutions, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," *IEEE Trans. Patt. Anal. Mach. Intell*, vol. 34, no. 7, pp. 1281–1298, 2011.

[18] M. F. Fallon, J. Folkesson, H. McClelland, and J. J. Leonard, "Relocating underwater features autonomously using sonar-based SLAM," *IEEE Journal of Oceanic Engineering*, vol. 38, no. 3, pp. 500–513, 2013.

[19] P. Ozog, G. Troni, M. Kaess, R. M. Eustice, and M. Johnson-Roberson, "Building 3d mosaics from an autonomous underwater vehicle, doppler velocity log, and 2d imaging sonar," in *Proc. IEEE Int. Conf. Robot. and Automation*, Seattle, WA, USA, 2015, pp. 1137–1143.

[20] J. Vaganay, L. Gurfinkel, M. Elkins, D. Jankins, and K. Shurn, "Hovering autonomous underwater vehicle-system design improvements and performance evaluation results," in *Proc. Int. Symp. Unmanned Untethered Subm. Tech.*, 2009, pp. 1–14.

[21] Sound Metrics Corp. (). Didson 300 m Imaging Sonar. Specification sheet and documentations Available at http://www.blueview.com/products/2d-imaging-sonar/.

[22] S. M. Chaves, A. Kim, and R. M. Eustice, "Opportunistic sampling-based planning for active visual SLAM," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots and Syst.*, Chicago, IL, USA, 2014, pp. 3073–3080.

[23] V. Ila, J. M. Porta, and J. Andrade-Cetto, "Information-based compact pose SLAM," *IEEE Transactions on Robotics*, vol. 26, no. 1, pp. 78–93, 2010.

[24] J. Li, P. Ozog, J. Abernethy, R. M. Eustice, and M. Johnson-Roberson, "Utilizing high-dimensional features for real-time robotic applications: Reducing the curse of dimensionality for recursive bayesian estimation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots and Syst.*, 2016, pp. 1230–1237.

[25] R. M. Eustice, O. Pizarro, and H. Singh, "Visually augmented navigation for autonomous underwater vehicles," *IEEE Journal of Oceanic Engineering*, vol. 33, no. 2, pp. 103–122, 2008.

[26] S. Agarwal, K. Mierle, and Others, *Ceres solver*, "http://ceres-solver.org".