SLAM for Ship Hull Inspection using Exactly Sparse Extended Information Filters

Matthew Walter, Franz Hover, and John Leonard
Department of Mechanical Engineering, Massachusetts Institute of Technology
{mwalter, hover, jleonard}@mit.edu

Abstract— **Many important missions for autonomous underwater vehicles (AUVs), such as undersea inspection of ship hulls, require integrated navigation, control, and motion planning in complex, 3D environments. This paper describes a SLAM implementation using forward-looking sonar (FLS) data from a highly maneuverable, hovering AUV performing a ship hull inspection mission. The Exactly Sparse Extended Information Filter (ESEIF) algorithm is applied to perform SLAM based upon features manually selected within FLS images. The results demonstrate the ability to effectively map a ship hull in a challenging marine environment. This provides a foundation for future work in which real-time SLAM will be integrated with motion planning and control to achieve autonomous coverage of a complete ship hull.**

I. Introduction

Frequent inspection of marine structures, including ships, walls, and jetties is a pressing need for governments and port authorities worldwide. There is a strong desire to remove humans from this task, instead developing autonomous agents are able to ensure complete coverage and the accurate localization and identification of anomalies on the structure. One of the challenges faced in this endeavor is that the environment characteristic of cluttered harbors provides little opportunity for traditional navigation: compasses are regularly corrupted by proximity to metal structures; the use of line-of-sight acoustic positioning systems are subject to limitations due to multipath returns, engine and equipment noise, and physical obstructions. Real-time, map-based navigation is thus an attractive means for performing the inspection based solely on sensors that interact with the local environment.

The requisite elements of an effective feature-based navigation system for this task include: a physical platform with low-level flight control that servos roll and pitch, such as that based on an IMU and depth sensor; sensors that provide information about the local environment, such as an imaging sonar and a Doppler velocity log (DVL); real-time algorithms that extract robust features from the sensor data streams; and an estimation framework that reconciles these feature observations with the other navigation information in order to jointly build a map and localize the vehicle within the map.

The present paper details full-scale experimental work that demonstrates 6 DOF Simultaneous Localization and Mapping (SLAM) using data from an imaging sonar mounted on a highly maneuverable AUV. This paper presents preliminary results in which sonar image features were detected manually and applied to a SLAM algorithm in post-processing in order to demonstrate the potential for automated ship hull inspection with forward-looking sonar, exploiting SLAM for ship-relative navigation.

Today's advanced sonar sensors produce acoustic images that look more and more like those of optical cameras. While the quality of optical images is superior in ideal conditions, sonars are far less sensitive to the limitations that characterize the underwater environment. In particular, optical cameras typically require external lighting to accommodate for the rapid attenuation of light with depth [1]. Additionally, the turbidity of the water severely limits the imaged field of view for standard cameras. The need for an external light source further compounds the effects of turbidity as it yields an increase in backscatter with larger fields of view [2]. Sonar imaging devices, on the other hand, rely upon the acoustic illumination of the scene and are not affected by the lack of natural light or the presence of particulates in the water. As harbors and coastal environments are particularly turbid, imaging sonars are the sensor of choice for shallow water operation.

One available imaging sonar is the Dual-Frequency Identification Sonar (DIDSON) [3], a high resolution sonar whose small size and low power consumption make it well-suited for operation on autonomous underwater vehicles. The DIDSON is a high frequency, forward-looking sonar that produces two-dimensional acoustic intensity images up to ranges of 40 meters at near-video rates. The sonar has a wide variety of applications that include monitoring fish populations [4], the inspection of underwater structures, explosive ordinance detection, and underwater mosaicing [5], [6]. These applications demonstrate the ability to achieve near optical quality imagery with an acoustic camera in poorly lit, turbid environments that are not conducive to optical cameras.

While the performance of acoustic cameras is impressive, the images lack the detail and accuracy of optical images taken under suitable conditions. Acoustic images suffer from a comparatively low signal-to-noise ratio (SNR), largely due to the effects of speckle noise in the sampled echos. Additionally, the resolution of DIDSON imagery is significantly lower than that of optical cameras. As a result of the acoustic beam pattern as well as physical constraints, the DIDSON is limited to a single linear receptor array of 96 transducers that each resolve intensity into 512 range bins. Contrast this with digital optical cameras that sample light with a two-dimensional array that contains millions of

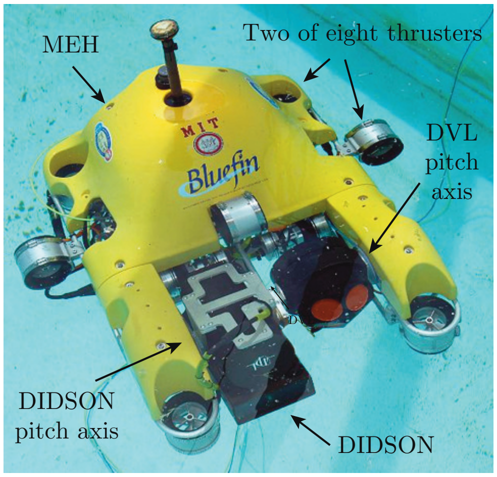

Fig. 1. The Hovering Autonomous Underwater Vehicle (HAUV) is roughly 90 cm (L) × 85 cm (W) × 45 cm (H) and weighs approximately 90 kg. The vehicle is equipped with an IMU, depth sensor, and compass that are housed within the main electronics housing (MEH). Located at the front of the vehicle, a DIDSON imaging sonar and DVL can be independently pitched to accommodate the geometry of the hull under inspection.

pixels. The relatively small number of transducer elements in the acoustic receiver array yields significantly reduced resolution.

The majority of camera-based pose and structure estimation algorithms rely upon low-level image interest points. For example, Eustice *et al.* [7] describe an algorithm that estimates the relative pose associated with overlapping images based upon a combination of shared SIFT [8] and Harris corner [9] features. The algorithm has been successfully applied for AUV navigation at depth where the visibility is sufficient with an external light source. The reduced resolution and lower SNR of acoustic imagery complicates the robust low-level feature detection and description necessary for such a delayed-state framework. Nonetheless, Kim *et al.* [5] utilize multiscale Harris corner interest points to identify feature correspondences between sets of DIDSON images. They then use the image correspondence to build a mosaic of the scene. Similarly, Negahdaripour *et al.* [6] describe an algorithm that generates DIDSON image mosaics based upon Harris corner features. Neither approach, though, estimates the three dimensional structure of the environment as they assume that the scene is planar. Additionally, the algorithms perform a batch optimization for the mosaic and and are not directly suitable for online localization and map building.

II. Autonomous Underwater Vehicle Platform

This paper presents underwater inspection work with the Hovering Autonomous Underwater Vehicle (HAUV). The HAUV was initially designed and built by MIT Sea Grant with the assistance of Bluefin Robotics as a platform for close-range underwater inspection [10], [11]. Figure 1 presents the most recent of the two versions of the HAUV [12].

The HAUV is equipped with navigation sensors as well as a forward-looking sonar. The navigation suite includes an IMU, which provides observations of the vehicle's three-axis angular rates as well as its pitch and roll angles. Heading is measured with a magnetic compass but, due to the interference induced by metal hulls, is unreliable during surveys. The vehicle measures its velocity and range relative to the ship's hull with a Doppler velocity log (DVL) that is steerable in pitch in order to face the hull. Over the course of the survey, the HAUV integrates this motion data to estimate its position with respect to the ship. These dead-reckoned estimates tend to be highly accurate over short timescales but give rise to errors in pose that grow over time.

Previous deployments of the HAUV have relied upon dead-reckoning to execute ship surveys [10]. Imagery collected by the vehicle is then reconstructed in the form of a mosaic as part of a post-processing step in order to view the entire area that was covered in the inspection. This process typically works only for relatively flat parts of the hull and yields only a partial model of the ship. A second limitation of this offline analysis, any gaps in the survey coverage can only be detected after post-processing, typically after the end of the mission. The aim of the present project is to use the acoustic imagery to perform localization and mapping concurrently over the course of the mission. This capability provides drift-free pose estimates, improves the accuracy of the map of the hull, and enables tasks such as the online assessment of coverage patterns.

III. Acoustic Imaging Sonar

The HAUV is equipped with a Dual-Frequency Identification Sonar (DIDSON) that records acoustic images of close-range underwater structures. The sonar provides near video rate (5 Hz-20 Hz) imagery of targets within a narrow field of view that extends to ranges upwards of 40 meters. Through a unique type of beamforming, the DIDSON samples the acoustic return intensity as a set of fixed-bearing temporal signals. By then sampling these profiles, the DIDSON generates a two-dimensional range versus bearing projection of the ensonification echo. Figure 2 shows the DIDSON image of a rectangular target on a ship hull taken at a range of roughly two meters. The range versus bearing image is transformed into Cartesian space, which demonstrates the DIDSON's non-rectangular field of view.

The DIDSON images targets within a narrow field of view that spans an angle of 28.8° in the azimuthal (bearing) direction and 12° in the vertical (elevation) direction. The range extent of the FOV depends upon the DIDSON operating frequency, which trades off lower resolution for increased range. At the lower 1.0 MHz frequency, the FOV extends 40 meters in front of the lens with a range resolution of 80 mm. Operating at 1.8 MHz, the maximum FOV range is 12 meters at a resolution of 20 mm. The HAUV typically conducts surveys with the DIDSON operating in the 1.8 MHz mode with a range window in the vicinity of 2.5 meters.

The DIDSON acquires acoustic imagery with a pair of lenses that focus a set of narrow ensonification beams at specific angles within the 28.8° azimuthal field of view. The beamwidth depends upon the operating mode of the

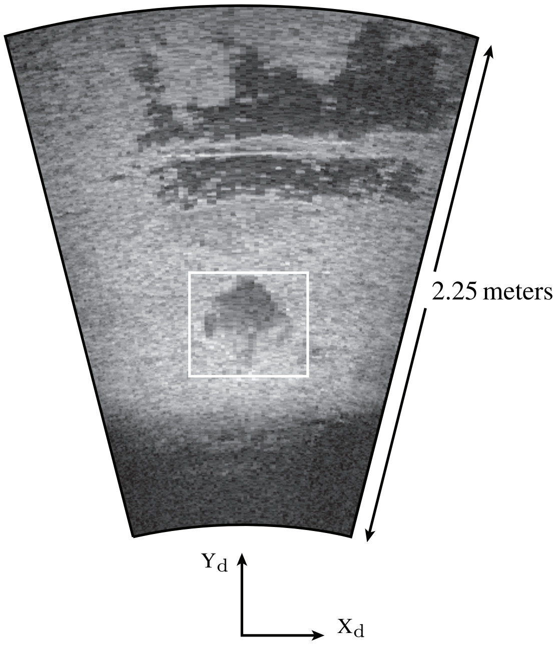

Fig. 2. The acoustic image of a rectangular target attached to a ship hull. The image represents the Cartesian projection of the range versus bearing intensity profile.

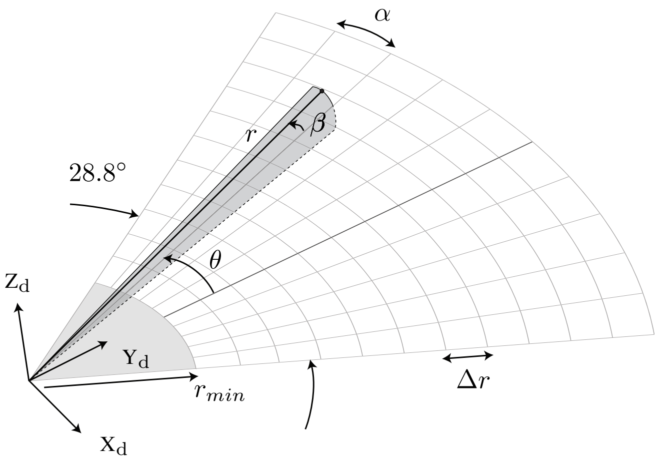

Fig. 3. Geometric representation of the DIDSON imaging model. The DIDSON produces a two-dimensional acoustic image that resolves intensity as a function of bearing, θ, and range, r, but not elevation, β.

DIDSON. In the lower frequency (1.0 MHz) extended range mode, there are 48 distinct beams that are each 0.6° wide. When operated at the higher frequency (1.8 MHz) identification mode, the DIDSON uses 96 beams, each of which are 0.3° wide. The vertical (elevation) beamwidth is 12° degrees in both cases. Using the same lenses, the DIDSON directs the acoustic returns onto a set of transducers that form a linear array. This "line-focused" beamforming yields an echo time profile for each of the 96 (48) beams. These profiles are then sampled to estimate range, resulting in 512 range bins along each beam direction. The result is a 512 by 96 (48) acoustic image that resolves return intensity into range and bearing. We depict this imaging geometry in Figure 3 where we define the location of imaged points in spherical coordinates, (r, θ, β), relative to a DIDSON-fixed coordinate frame.

The DIDSON does not disambiguate the elevation of the target and, as a result, echos may originate from anywhere along a $|\beta| \leq 12^\circ$ arc. An alternative to the constant-radius projection is to approximate this mapping with a model that projects pixels along lines that are perpendicular to the image plane. Due to the narrow FOV in elevation ($|\beta| \leq 6^\circ$), this approximation results in relatively little error, approximately 0.5% of the range to the target. Kim *et. al* [5] show that this projection approximation gives rise to an orthographic camera model for the DIDSON. Considering two overlapping images taken from different viewpoints, this model results in an affine transformation that relates their corresponding camera frames [13]. Kim *et. al* exploit the epipolar geometry to estimate the affine transformation between pairs of images and, in turn, generate mosaics.

The epipolar geometry provides constraints on the relative transformation between pairs of vehicle poses. These constraints can be integrated as part of a SLAM filter. The data is particularly amenable to a pose graph framework that performs estimation over a state consisting of the robot position history based upon pairwise constraints between poses. For example, Eustice *et al.* [7] rely upon this data to track a probability distribution over the history of vehicle camera poses. The algorithm treats the epipolar constraints between camera pairs as measurements of the corresponding vehicle poses and incorporates the observations in an Extended Kalman Filter (EKF) step. This and other pose graph techniques offer benefits related to performance, computation time, and robustness [14]–[16].

The epipolar geometry that governs optical cameras gives rise to five constraints on relative camera pose. In the case of orthographic camera models, though, the affine epipolar geometry is invariant to a greater degree of relative motion. This invariance reduces the number of constraints that are resolvable from pairs of overlapping images to only three [13]. For that reason, we instead take an online, feature-based approach to SLAM[1] whereby we maintain an estimate of the current robot pose along with a set of landmarks that comprise the map. In the context of ship hull inspection, these targets include both natural features as well as man-made objects. We currently identify these targets by hand-selecting features within the DIDSON imagery, though the estimation framework is amenable to the automated detection of interest regions (i.e. greater than pixel resolution). In the next section, we describe a novel SLAM algorithm that exploits these features to accurately localize the vehicle in a consistent and efficient manner.

IV. Feature-based SLAM Based Upon Acoustic Imagery

Underwater survey applications pose the problem of generating a model of the structure as one of map building.

[1] Here, "online" SLAM refers to the fact that we are only concerned with the current robot pose and not its trajectory history.

This framework requires accurate knowledge of vehicle pose, typically estimated with the help of an acoustic long baseline (LBL) beacon network, which provides triangulated time-of-flight position data. Unfortunately, LBL and similar variants are less than optimal for the in-situ inspection of ship hulls as they typically rely upon the initial deployment and calibration of the beacon infrastructure. Additionally, the environment in which surveys generally take place complicate time-of-flight localization as a result of the multipath interference that results from the shallow water column and the vehicle's close proximity to the ship hull.

Rather than rely on external infrastructure for localization, we view the ship hull inspection problem in the Simultaneous Localization and Mapping (SLAM) framework [17]. SLAM seeks to build a map of the environment based upon the vehicle pose estimate and to concurrently localize the robot within this map. This coupling is complicated by the fact that both the vehicle motion as well as the observations of the environment are prone to uncertainty. The majority of SLAM algorithms address these issues by posing the problem in a probabilistic framework whereby they track a joint distribution over the vehicle pose and map.

A. *Feature-based SLAM*

We adopt a feature-based representation for SLAM and model the 3D ship hull environment as a collection of object primitives, $\mathbf{M} = \{\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_n\}$. The map, together with the current robot pose comprise the state vector at time t, $\boldsymbol{\xi}_t = [\mathbf{x}_t^\top \;\; \mathbf{M}^\top]^\top$. We define the 6-DOF vehicle state, $\mathbf{x}_t$, by a 12 element vector that includes the vehicle's position, its roll, pitch, and heading Euler angles, as well as the vehicle's linear and angular velocity in the body-fixed frame. We wish to track the joint distribution over the vehicle state and map, $p\left(\boldsymbol{\xi}_t|\mathbf{z}^t, \mathbf{u}^t\right)$, based upon the entire history of measurement and motion data, $\mathbf{z}^t = \{z_1, z_2, \ldots, z_t\}$ and $\mathbf{u}^t = \{u_1, u_2, \ldots, u_t\}$, respectively.

We model the uncertainty in the HAUV motion model along with the DIDSON observations of the ship hull as Gaussian white noise. Linearizing the process and measurement models, we are then able to represent the posterior likelihood by a Gaussian distribution,

$$p\left(\boldsymbol{\xi}_t|\mathbf{z}^t, \mathbf{u}^t\right) = \mathcal{N}\big(\boldsymbol{\xi}_t; \boldsymbol{\mu}_t, \Sigma_t\big) = \mathcal{N}^{-1}\big(\boldsymbol{\xi}_t; \boldsymbol{\eta}_t, \Lambda_t\big) \qquad (1)$$

The *standard* form of the Gaussian parametrizes the distribution in terms of the mean vector, $\boldsymbol{\mu}_t$, and covariance matrix, Σ_t. Alternatively, the *canonical* form represents the Gaussian by the information matrix, Λ_t, and information vector, $\boldsymbol{\eta}_t$. The two forms are duals of one another, related by the relationship:

$$\Lambda_t = \Sigma_t^{-1} \quad \boldsymbol{\eta}_t = \Lambda_t \boldsymbol{\mu}_t \qquad (2)$$

Beginning with the seminal work Smith, Self, and Cheeseman [18], the Extended Kalman Filter has laid the groundwork for numerous successful SLAM algorithms that track the standard form of the posterior. One property of the standard form that is both a benefit and a burden is that it explicitly maintains the coupling between the robot and map in the form of a dense covariance matrix. This leads to the well-known computational and memory costs that are quadratic in the size of the map, limiting the application of the EKF to environments with hundreds of features [17].

The canonical parametrization of the Gaussian has been proposed as a solution to the feature-based SLAM scalability problem. Pivotal insights by Thrun *et al.* [19] and Frese *et al.* [20] reveal that, unlike the covariance matrix, its inverse information matrix is *nearly* sparse with a majority of the elements near zero. In the case that the parametrization can be approximated as being truly sparse, Thrun *et al.* [19], Frese *et al.* [20], and Paskin [21] each present SLAM algorithms with complexity that is near-constant time.[2] Most directly related to the work described here, the Sparse Extended Information Filter (SEIF) [19] is a variation on the Extended Information Filter (EIF) [23], the dual of the EKF. The SEIF periodically sparsifies the information matrix by ignoring the conditional dependence relationships between the vehicle and a subset of the map. Once sparse, the SEIF implements the EIF time projection and measurement update steps in constant time given an estimate for the mean. Eustice, Walter, and Leonard [24], though, demonstrate that the SEIF sparsification strategy, which ignores conditional dependencies, results in overconfident state estimates. The consequences include a resulting distribution over the map that is inconsistent along with the inability to correctly match observations with mapped features (data association) [25].

B. *Exactly Sparse Extended Information Filter*

Walter *et al.* [26] present the Exactly Sparse Extended Information Filter (ESEIF) as an alternative sparse information filter that achieves the computational benefits of a sparse parametrization while preserving consistency. As with the SEIF, the ESEIF is a modified version of the EIF that maintains a sparse information matrix. The key contribution of the ESEIF algorithm is a sparsification strategy that maintains an information matrix in which the majority of elements are *exactly* zero. In turn, the ESEIF avoids the need to approximate conditional independencies and thereby preserves the consistency of the Gaussian distribution. The ESEIF then maintains map and pose estimates that are nearly identical to those of the EKF but exploits the sparse SLAM parametrization to track the distribution in near-constant time. The following provides a brief introduction to the prediction and update steps as implemented in the ESEIF. For a more detailed discussion, the reader is referred to the description of Walter, Eustice, and Leonard [27].

1) Measurement Update Step: Throughout the survey of the hull, the vehicle makes noisy observations of the state vector. Most importantly, these include measurements of the range and bearing to mapped targets on the ship hull provided by the DIDSON imagery. We also treat proprioceptive observations of depth, attitude, angular rates, and hull-relative velocity as measurements of vehicle state. In

[2] The particular mean recovery process dictates the computational cost. Partial mean recovery is typically constant-time but occasionally requires full mean estimation, which is generally linear in the size of the map [22].

the most general form, we model observations as nonlinear functions of the state that are corrupted by white Gaussian noise, $\mathbf{v}_t \sim \mathcal{N}\big(\mathbf{0}, \mathrm{R}\big)$. Equation (3b) is the linearization about the current mean for the observed state elements and H the corresponding sparse Jacobian.

$$\begin{aligned} \mathbf{z}_t &= \mathbf{h}\big(\boldsymbol{\xi}_t\big) + \mathbf{v}_t && (3a)\\ &\approx \mathbf{h}\big(\bar{\boldsymbol{\mu}}_t\big) + \mathrm{H}\big(\boldsymbol{\xi}_t - \bar{\boldsymbol{\mu}}_t\big) + \mathbf{v}_t && (3b) \end{aligned}$$

The filter incorporates the measurement information in the SLAM posterior via the measurement update step, which involves the addition of a sparse matrix to the current information matrix:

$$\begin{aligned} p\left(\boldsymbol{\xi}_t \mid \mathbf{z}^t, \mathbf{u}^t\right) &= \mathcal{N}^{-1}\big(\boldsymbol{\eta}_t, \Lambda_t\big) \\ \Lambda_t &= \bar{\Lambda}_t + \mathrm{H}^\top \mathrm{R}^{-1}\mathrm{H} && (4a)\\ \boldsymbol{\eta}_t &= \bar{\boldsymbol{\eta}}_t + \mathrm{H}^\top \mathrm{R}^{-1}\big(\mathbf{z}_t - \mathbf{h}\left(\bar{\boldsymbol{\mu}}_t\right) + \mathrm{H}\bar{\boldsymbol{\mu}}_t\big) && (4b) \end{aligned}$$

At any point in time, the HAUV will observe only a small number of targets on the hull due to the DIDSON's limited FOV. Including these measurements along with those of the vehicle's attitude and body velocities, the Jacobian, H, is sparse with a bounded number of non-zero elements. The computational cost of the matrix product, $\mathrm{H}^\top\mathrm{R}^{-1}\mathrm{H}$, within (4a) is constant-time and the addition only modifies entries of the information matrix that exclusively correspond to observed states. Assuming that we have an estimate of the mean for the vehicle pose and the imaged landmarks, the canonical update in (4) is quadratic in the number of measurements, irrespective of the sparsity of the information matrix.

2) Time Prediction Step: The vehicle state vector evolves over time according to a constant velocity model of the kinematics (5a) that is nonlinear in the state and control input. An additive, white Gaussian noise term, $\mathbf{w}_t \sim \mathcal{N}\big(\mathbf{0}, \mathrm{Q}\big)$, captures uncertainty in the dynamics model that we assume to be first-order Markov. The linearization of the process model follows in (5b) where F denotes the Jacobian matrix.

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{f}\big(\mathbf{x}_t, \mathbf{u}_{t+1}\big) + \mathbf{w}_t && (5a)\\ &\approx \mathbf{f}\big(\boldsymbol{\mu}_{x_t}, \mathbf{u}_{t+1}\big) + \mathrm{F}\big(\mathbf{x}_t - \boldsymbol{\mu}_{x_t}\big) + \mathbf{w}_t && (5b) \end{aligned}$$

We update the distribution to reflect the robot motion by first augmenting the state with the new vehicle pose and subsequently marginalize over the previous pose. The combined effect yields the time prediction step for the ESEIF:

$$\begin{aligned} p\left(\boldsymbol{\xi}_{t+1} \mid \mathbf{z}^t, \mathbf{u}^{t+1}\right) &= \mathcal{N}^{-1}\big(\bar{\boldsymbol{\eta}}_{t+1}, \bar{\Lambda}_{t+1}\big) \\ \bar{\boldsymbol{\eta}}_{t+1} &= \begin{bmatrix} \mathrm{Q}^{-1}\Delta \\ \bar{\boldsymbol{\eta}}_M \end{bmatrix} - \begin{bmatrix} -\mathrm{Q}^{-1}\mathrm{F} \\ \Lambda_{Mx_t} \end{bmatrix} \Omega^{-1}\big(\boldsymbol{\eta}_{x_t} - \mathrm{F}^\top\mathrm{Q}^{-1}\Delta\big) \\ &= \begin{bmatrix} \mathrm{Q}^{-1}\mathrm{F}\Omega^{-1}\boldsymbol{\eta}_{x_t} + \Psi\Delta \\ \boldsymbol{\eta}_M - \Lambda_{Mx_t}\Omega^{-1}\big(\boldsymbol{\eta}_{x_t} - \mathrm{F}^\top\mathrm{Q}^{-1}\Delta\big) \end{bmatrix} && (6a) \end{aligned}$$

$$\begin{aligned} \bar{\Lambda}_{t+1} &= \begin{bmatrix} \mathrm{Q}^{-1} & 0 \\ 0 & \Lambda_{MM} \end{bmatrix} - \begin{bmatrix} \mathrm{Q}^{-1}\mathrm{F} \\ \Lambda_{Mx_t} \end{bmatrix} \Omega^{-1} \begin{bmatrix} -\mathrm{F}^\top\mathrm{Q}^{-1} & \Lambda_{x_tM} \end{bmatrix} \\ &= \begin{bmatrix} \Psi & \mathrm{Q}^{-1}\mathrm{F}\Omega^{-1}\Lambda_{x_tM} \\ \Lambda_{Mx_t}\Omega^{-1}\mathrm{F}^\top\mathrm{Q}^{-1} & \Lambda_{MM} - \Lambda_{Mx_t}\Omega^{-1}\Lambda_{x_tM} \end{bmatrix} && (6b) \end{aligned}$$

where

$$\begin{aligned} \Delta &= \mathbf{f}\left(\mathbf{x}_t, \mathbf{u}_{t+1}\right) - \mathrm{F}\boldsymbol{\mu}_{x_t} \\ \Psi &= \mathrm{Q}^{-1} - \mathrm{Q}^{-1}\mathrm{F}\left(\Lambda_{x_tx_t} + \mathrm{F}^\top\mathrm{Q}^{-1}\mathrm{F}\right)^{-1}\mathrm{F}^\top\mathrm{Q}^{-1} \\ &= \left(\mathrm{Q} + \mathrm{F}\Lambda_{x_tx_t}^{-1}\mathrm{F}^\top\right)^{-1} \\ \Omega &= \Lambda_{x_tx_t} + \mathrm{F}^\top\mathrm{Q}^{-1}\mathrm{F} \end{aligned}$$

Note that the time prediction step induces the fill-in of the information matrix. This is evident in the map sub-block of the new information matrix (6b), $\Lambda_{MM} - \Lambda_{Mx_t}\Omega^{-1}\Lambda_{x_tM}$, in which the matrix $\Lambda_{Mx_t} = \Lambda_{x_tM}^\top$ is non-zero for any *active* map element that shares information with the robot pose. The matrix outer product, $\Lambda_{Mx_t}\Omega^{-1}\Lambda_{x_tM}$, then creates shared information between the set of active landmarks. Since features are forever active once they are added to the map, the information matrix becomes fully dense.

Furthermore, this matrix product in (6b) is the most computationally expensive aspect of the prediction step. The number of multiplication operations necessary to compute this outer product is quadratic in the number of active features. Since each landmark will inherently be active, this implies that, in its natural, non-sparse form, the EIF prediction step is quadratic in the size of the map.

3) ESEIF Sparsification: A limit on the active map size plays a critical role in improving the computational and memory performance of feature-based SLAM information filters. For one, a bound on the number of off-diagonal entries in the information matrix that link the robot pose and map limits the number of non-zero elements that are added as a consequence of time prediction. Secondly, an active map bound allows us to control the computational cost of the prediction step, irrespective of the size of the map.

The ESEIF takes a proactive approach to sparsifying the information matrix whereby it controls the initial formation of elements in the matrix. The filter essentially performs periodic preventative maintenance to regulate the fill-in of the information matrix. The general idea behind sparsification is relatively straightforward. Rather than allow the matrix to grow dense and occasionally prune weak elements to preserve sparsity, the ESEIF manages the initial formation of shared information between the robot and map. The sparsification strategy does this in a novel way, by periodically "kidnapping" (marginalizing) the robot from the map and subsequently relocalizing it based upon a limited number of measurements. The only links that result between the map and the vehicle state correspond to the features used for relocalization. The filter then continues as before, repeating the same marginalization and relocalization process when necessary.

The actual sparsification procedure takes the form of a modified measurement update step. The ESEIF partitions the current set of observations into two subsets, $\mathbf{z}_\alpha$ and $\mathbf{z}_\beta$. In the first step of sparsification, the filter updates the SLAM distribution, $p\left(\boldsymbol{\xi}_t \mid \mathbf{z}^{t-1}, \mathbf{u}^t\right) = \mathcal{N}^{-1}\big(\boldsymbol{\eta}_t, \Lambda_t\big)$, as in (4) based upon the observations in $\mathbf{z}_\alpha$:

$$p_1\big(\boldsymbol{\xi}_t \mid \left\{\mathbf{z}^{t-1}, \mathbf{z}_\alpha\right\}, \mathbf{u}^t\big) = \mathcal{N}^{-1}\big(\bar{\boldsymbol{\eta}}_t, \bar{\Lambda}_t\big).$$

The ESEIF subsequently marginalizes out the robot pose from the posterior to achieve a distribution over only the map,

$$\begin{aligned} p_2\big(\mathbf{M} \mid \{\mathbf{z}^{t-1},\mathbf{z}_\alpha\},\mathbf{u}^t\big) &= \int_{\mathbf{x}_t} p_1\big(\boldsymbol{\xi}_t \mid \{\mathbf{z}^{t-1},\mathbf{z}_\alpha\},\mathbf{u}^t\big)d\mathbf{x}_t \\ &= \mathcal{N}^{-1}\big(\mathbf{M};\breve{\boldsymbol{\eta}}_t,\breve{\Lambda}_t\big). \end{aligned} \tag{7}$$

The corresponding information matrix and information vector for the map distribution follow as,

$$p_2\big(\mathbf{M}_t \mid \{\mathbf{z}^{t-1},\mathbf{z}_\alpha\},\mathbf{u}^t\big) = \mathcal{N}^{-1}\big(\mathbf{M}_t;\breve{\boldsymbol{\eta}}_t,\breve{\Lambda}_t\big)$$

$$\breve{\Lambda}_t = \bar{\Lambda}_{MM} - \bar{\Lambda}_{Mx_t}\left(\bar{\Lambda}_{x_tx_t}\right)^{-1}\bar{\Lambda}_{x_tM} \tag{8a}$$

$$\breve{\boldsymbol{\eta}}_t = \bar{\boldsymbol{\eta}}_M - \bar{\Lambda}_{Mx_t}\left(\bar{\Lambda}_{x_tx_t}\right)^{-1}\bar{\boldsymbol{\eta}}_{x_t}. \tag{8b}$$

The matrix outer product in (8a) dictates the computational complexity of this step, which is quadratic in the number of non-zero terms between the robot state and map in the information matrix. Since we are bounding this number, this marginalization is constant-time.

The final component of ESEIF sparsification utilizes the remaining $\mathbf{z}_\beta$ measurements to relocalize the robot within the map. We typically appropriate the $\mathbf{z}_\beta$ set by reserving as many DIDSON observations as possible for relocalization along with the direct measurements of the HAUV attitude, depth, and body velocities. In general, the relocalized vehicle state is a nonlinear function (9a) of $\mathbf{z}_\beta$ and the corresponding map elements, $\mathbf{m}_\beta$. We represent error in this model with Gaussian noise, $\mathbf{w}_t \sim \mathcal{N}\big(\mathbf{0},\mathrm{R}\big)$. The linearization about the mean follows in (9b) where the Jacobian, G, is non-zero only at positions that correspond with $\mathbf{m}_\beta$.

$$\mathbf{x}_t = \mathbf{g}\big(\mathbf{m}_\beta,\mathbf{z}_\beta\big) + \mathbf{w}_t \tag{9a}$$

$$\approx \mathbf{g}\big(\breve{\boldsymbol{\mu}}_{m_\beta},\mathbf{z}_\beta\big) + \mathrm{G}\big(\mathbf{m}-\breve{\boldsymbol{\mu}}_t\big) + \mathbf{w}_t \tag{9b}$$

The implementation of the relocalization component of sparsification is essentially no different from the process of adding new features to the map. It is then straightforward to show that the canonical form of the SLAM posterior follows as

$$p_{\text{ESEIF}}\big(\boldsymbol{\xi}_t \mid \mathbf{z}^t,\mathbf{u}^t\big) = \mathcal{N}^{-1}\big(\boldsymbol{\xi}_t;\breve{\boldsymbol{\eta}}_t,\breve{\Lambda}_t\big)$$

$$\breve{\Lambda}_t = \begin{bmatrix} \mathrm{R}^{-1} & -\mathrm{R}^{-1}\mathrm{G} \\ -\mathrm{G}^\top\mathrm{R}^{-1} & \breve{\Lambda}_t + \mathrm{G}^\top\mathrm{R}^{-1}\mathrm{G} \end{bmatrix} \tag{10a}$$

$$\breve{\boldsymbol{\eta}}_t = \begin{bmatrix} \mathrm{R}^{-1}\left(\mathbf{g}\big(\breve{\boldsymbol{\mu}}_{m_\beta},\mathbf{z}_\beta\big) - \mathrm{G}\breve{\boldsymbol{\mu}}_t\right) \\ \breve{\boldsymbol{\eta}}_t - \mathrm{G}^\top\mathrm{R}^{-1}\left(\mathbf{g}\big(\breve{\boldsymbol{\mu}}_{m_\beta},\mathbf{z}_\beta\big) - \mathrm{G}\breve{\boldsymbol{\mu}}_t\right) \end{bmatrix} \tag{10b}$$

As a consequence of the sparsity of G, the modification to the map sub-block of the information matrix in (10a), $\breve{\Lambda}_t + \mathrm{G}^\top\mathrm{R}^{-1}\mathrm{G}$ is constant-time and affects only the $\mathbf{m}_\beta$ elements. Furthermore, the off-diagonal term, $\mathrm{R}^{-1}\mathrm{G}$, is zero everywhere except for entries associated with the $\mathbf{m}_\beta$ features. By periodically "kidnapping" the robot from the map and relocalizing it, the ESEIF sparsification strategy preserves an exactly sparse information matrix at a slight cost of lost temporal information between poses [26], [27].

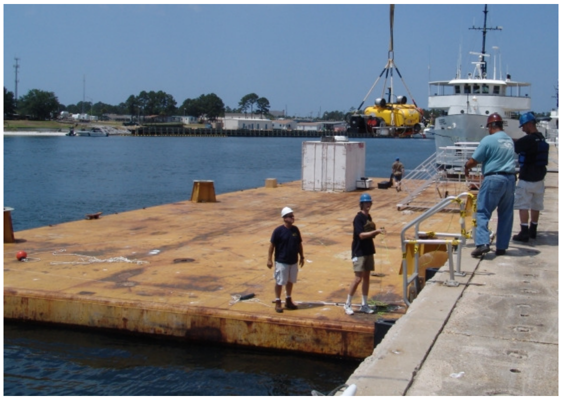

Fig. 4. A photograph of the first version of the HAUV being lowered into the water to survey the barge during AUVFest. The barge is 13.4 m from port (side nearest to the pier) to starboard and 36.2 m from bow to stern.

V. Results

We apply the ESEIF localization and mapping architecture to survey a series of man-made and natural targets located on a ship hull. The deployment that we consider was part of the 2007 AUVFest at the Naval Surface Warfare Center in Panama City, FL. The focus of the ship hull inspection experiments was a barge of length 36.2 meters and width 13.4 meters, shown in Figure 4, that was moored to a pier on its port side. Approximately 30 targets were distributed over the underside of the hull and their position measured by a team of divers. Among these objects were box targets of the form imaged in Figure 2, cylindrical "cake" targets roughly 20 cm in diameter, and over a dozen small brick-shaped objects. In addition to these features, the hull was littered with both man-made as well as natural targets, most of which are clearly visible in the DIDSON imagery.

The two HAUV vehicles spent more than thirteen hours over the course of the experiments collecting high resolution imagery of the barge. We consider a 45 minute survey of most of the barge that consists of four overlapping surveys of the bow, stern, port, and starboard sections of the hull. The vehicle starts the mission near the aft-starboard corner of the barge and first surveys most of the stern with the exception of the corners. The vehicle then proceeds to image the port and starboard sides, followed by the bow. The HAUV moves laterally along tracklines that span the width (for the stern and bow surveys) and length (for the starboard and port surveys) of the barge at a velocity of 25 cm/s. Throughout the survey, the DVL is positioned vertically upwards at the hull and the DIDSON is oriented at just over 20° from horizontal to achieve a suitable grazing angle with the hull. Over the duration of the nearly 45 minute mission, the HAUV collected about 4200 DIDSON images of the bottom of the barge.

In order to analyze the performance of the ESEIF algorithm, we hand-select features within the DIDSON imagery, in lieu of an automated detection process. Each detection provides a measure of the relative range and bearing to a target on the hull that is subject to the DIDSON's 12°

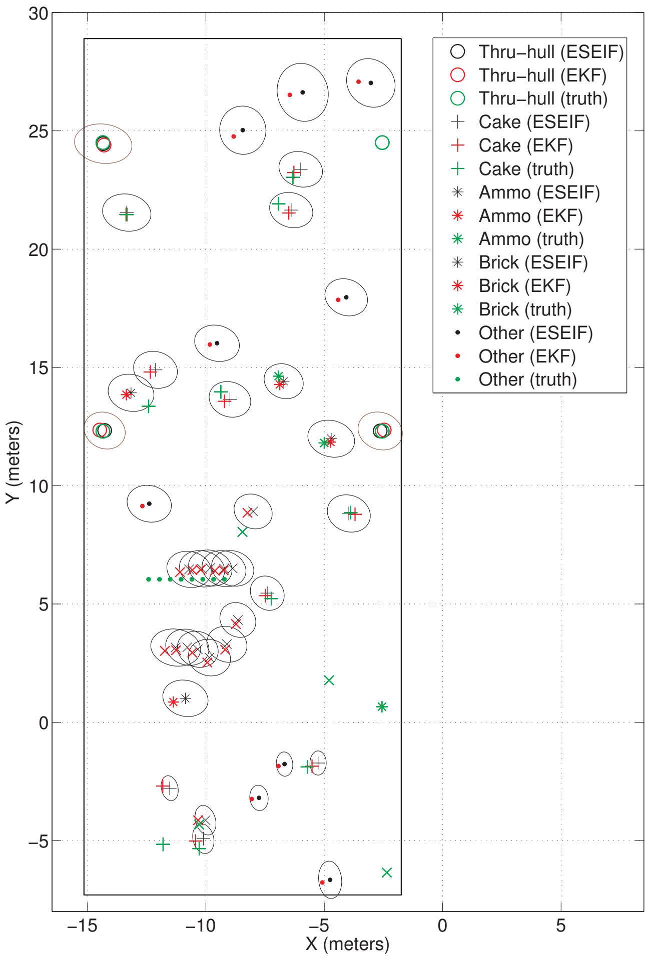

Fig. 5. Overhead view of the ESEIF map of the barge based upon hand-picked image features. The plot includes the EKF estimates for the feature locations as well as a measure of ground truth. Targets shown in black comprise the ESEIF map while the EKF map is shown in red and the ground truth in green. The ellipses denote the three-sigma uncertainty bounds maintained by the ESEIF.

elevation ambiguity. We resolve the ambiguity in elevation by independently tracking the local geometry of the hull based upon the range data from the DVL. We do not rely on any *a priori* information regarding the shape of the hull and only assume that it can be approximated as locally planar. The resulting measurement data serve as observations of the environment for the ESEIF algorithm.

We implement the ESEIF feature-based SLAM framework, utilizing the sparsification strategy to maintain a bound of five active landmarks. When sparsification is necessary, we partition the measurement data, $\mathbf{z}_t = \{\mathbf{z}_\alpha, \mathbf{z}_\beta\}$, such that relocalization utilizes all available measurements, i.e. $\mathbf{z}_\beta = \mathbf{z}_t$ and $\mathbf{z}_\alpha = \{\}$. As a basis for comparison, we concurrently apply the localization and mapping algorithm with the standard feature-based EKF estimator in place of the ESEIF.

Figure 5 presents a bird's eye view of the final SLAM maps. The plot compares the map built with the ESEIF with that of the "gold standard" EKF as well as the ground truth

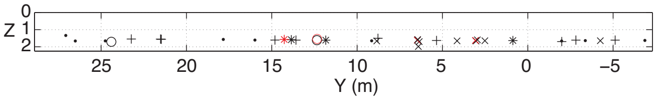

Fig. 6. A side view of the ESEIF map from the barge's port side. Note that the plot renders features with two colors to help discern between targets that overlap under this view. While there is no ground truth data regarding the depth of the targets, the DVL ranges to the hull suggest a uniform hull draft of 1.5 m. The mean feature depth as estimated by the filter is 1.63 m with a variance of 8.7 cm. The variation from our DVL-based estimate of the barge's draft is largely due to the three-dimensional structure of the targets, which we model as point features.

target locations as measured by the divers. Both the ESEIF and EKF maps are aligned with the barge thru-hulls based upon a least squares estimate for the transformation. The uncertainty ellipses correspond to the three-sigma confidence bounds associated with the ESEIF map estimates. Note that these intervals capture each of the EKF target positions, but not the ground truth location of every feature. We find the same to be true of EKF-based map estimates and believe that the disagreement is largely due to the divers' difficulty in accurately measuring the true position of the targets on the hull. Additionally, the ground truth data indicates that there are several targets that neither the EKF nor the ESEIF-based algorithms incorporate into their respective maps. An inspection of the images that are associated with these regions of the hull according to the ESEIF pose estimates suggests that these features broke free from hull. While this does not offer conclusive proof, it is an agreement with divers' claims that targets had broken free.

Meanwhile, we assess the 3D quality of the map based upon the depth of the features. Figure 6 presents a side view of the ESEIF map from the barge's port side. Ground truth data regarding the draft profile of the barge is unavailable. Based upon the vehicle's depth measurements and the DVL ranges to the hull, we estimate the draft to be 1.5 m. In comparison, the mapped features exhibit a mean depth of 1.63 m and a standard deviation of 8.7 cm. The synthetic targets are not flush with the hull and their vertical extent largely accounts for this offset.

In order to confirm that the ESEIF sparsification strategy does not induce overconfidence in the state estimates, we compare resulting uncertainty with that of the EKF. Specifically, we compute the ratio between the determinant of each feature's sub-block of the covariance (inverse information) matrix as maintained by the ESEIF with that of the EKF. On a log scale, a ratio greater than zero implies a conservative estimate for the uncertainty with respect to the EKF while negative ratios suggest overconfidence. We plot a histogram over these ratios in Figure 7. As we have found both in simulation as well as with experimental data [26], the plot confirms that the ESEIF preserves consistency relative to the EKF.

VI. Conclusion

This paper describes the application of the ESEIF algorithm to ship hull inspection using an AUV equipped with a

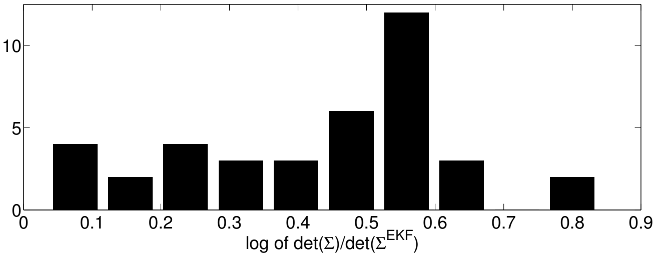

Fig. 7. A histogram plot comparing the ratio of feature uncertainty as estimated by the ESEIF with that of the EKF. Ratios greater than zero are indicative of conservative confidence intervals while negative values indicate overconfidence.

forward-looking sonar. The ESEIF provides a computationally efficient filtering framework that is amenable to large-scale joint estimation of the vehicle pose (coverage) history and the map of the hull. The results demonstrate the ability to produce maps that accurately identify the position of targets on the ship hull. The work that we have presented relies upon the manual selection of features within the acoustic imagery. We have similarly applied our framework with features detected automatically within FLS imagery, though the current detector suffers from low repeatability which degrades the performance of the filter. We are currently working on improving the performance of the automated detection process. Meanwhile, the goal of future work is to integrate the SLAM capability demonstrated here with real-time trajectory planning and control of the HAUV in order to achieve 100% coverage of an unknown ship.

VII. Acknowledgements

The authors would like to thank Mike Kokko and Dan Walker from MIT as well as Jerome Vaganay and Steve Summit at Bluefin Robotics for their help in conducting experiments both in preparation for and as part of AUVFest. This research has been supported by ONR grant N000140610043.

References

[1] H. Singh, C. Roman, O. Pizarro, and R. Eustice, "Advances in high resolution imaging from underwater vehicles," in *Proceedings of the 12th International Symposium of Robotics Research (ISRR)*, S. Thrun, R. Brooks, and H. Durrant-Whyte, Eds. San Francisco, CA: Springer, October 2005, pp. 430–448.

[2] J. Jaffe, "Computer modeling and the design of optimum underwater imaging systems," *IEEE Journal of Oceanic Engineering*, vol. 15, no. 2, pp. 1001–111, April 1990.

[3] E. Belcher, B. Matsuyama, and G. Trimble, "Object identification with acoustic lenses," in *Proceedings of OCEANS MTS/IEEE Conference and Exhibition*, vol. 1, Honolulu, HI, November 2001, pp. 6–11.

[4] J. Holmes, G. Cronkite, H. Enzenhofer, and T. Mulligan, "Accuracy and precision of fish-count data from a "dual-frequency identification sonar" (DIDSON) imaging system," *ICES Journal of Marine Science*, vol. 63, no. 3, pp. 543–555, 2006.

[5] K. Kim, N. Intrator, and N. Neretti, "Image registration and mosaicing of acoustic camera images," in *Proceedings of the 4th IASTED International Conference on Visualization, Imaging, and Image Processing*, 2004, pp. 713–718.

[6] S. Negahdaripour, P. Firoozfam, and P. Sabzmeydani, "On processing and registration of forward-scan acoustic video imagery," in *Proceedings of the Second Canadian Conference on Computer and Robot Vision*, May 2005, pp. 452–459.

[7] R. Eustice, O. Pizarro, and H. Singh, "Visually augmented navigation in an unstructured environment using a delayed state history," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, vol. 1, New Orleans, LA, 2004, pp. 25–32.

[8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, November 2004.

[9] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of the 4th Alvey Vision Conference*, Manchester, U.K, 1988, pp. 147–151.

[10] J. Vaganay, M. L. Elkins, S. Willcox, F. S. Hover, R. S. Damus, S. Desset, J. P. Morash, and V. C. Polidoro, "Ship hull inspection by hull-relative navigation and control," in *Proceedings of OCEANS MTS/IEEE Conference and Exhibition*, Washington, DC, 2005, pp. 761–766.

[11] 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, 2007.

[12] M. A. Kokko, "Range-based navigation of AUVs operating near ship hulls," Master's thesis, Massachusetts Institute of Technology, June 2007.

[13] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.

[14] R. Eustice, H. Singh, and J. Leonard, "Exactly sparse delayed-state filters," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Barcelona, Spain, April 2005, pp. 2417–2424.

[15] F. Dellaert, "Square root SAM," in *Proceedings of Robotics: Science and Systems (RSS)*, Cambridge, MA, June 2005, pp. 177–184.

[16] E. Olson, J. Leonard, and S. Teller, "Fast iterative optimization of pose graphs with poor initial estimates," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, May 2006, pp. 2262–2269.

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

[18] R. Smith, M. Self, and P. Cheeseman, "Estimating uncertain spatial relationships in robotics," in *Autonomous Robot Vehicles*, I. Cox and G. Wilfong, Eds. Springer-Verlag, 1990, pp. 167–193.

[19] S. Thrun, D. Koller, Z. Ghahramani, H. Durrant-Whyte, and A. Ng, "Simultaneous mapping and localization with sparse extended information filters: Theory and initial results," in *Proceedings of the Fifth International Workshop on Algorithmic Foundations of Robotics*, Nice, France, 2002.

[20] U. Frese and G. Hirzinger, "Simultaneous localization and mapping - a discussion," in *Proceedings of the IJCAI Workshop on Reasoning with Uncertainty in Robotics*, 2001, pp. 17–26.

[21] M. Paskin, "Thin junction tree filters for simultaneous localization and mapping," University of California, Berkeley, Tech. Rep. UCB/CSD-02-1198, September 2002.

[22] R. Eustice, "Large-area visually augmented navigation for autonomous underwater vehicles," Ph.D. dissertation, Massachusetts Institute of Technology / Woods Hole Oceanographic Institution Joint Program, Cambridge, MA, June 2005.

[23] A. G. Mutambara, *Decentralized Estimation and Control for Multisensor Systems*. Boston, MA: CRC Press, 1998.

[24] R. Eustice, M. Walter, and J. Leonard, "Sparse extended information filters: Insights into sparsification," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Edmonton, Alberta, Canada, August 2005, pp. 641–648.

[25] R. Eustice, H. Singh, J. Leonard, M. Walter, and R. Ballard, "Visually navigating the RMS Titanic with SLAM information filters," in *Proceedings of Robotics: Science and Systems (RSS)*, Cambridge, MA, June 2005, pp. 57–64.

[26] M. Walter, R. Eustice, and J. Leonard, "A provably consistent method for imposing exact sparsity in feature-based SLAM information filters," in *Proceedings of the 12th International Symposium of Robotics Research (ISRR)*, S. Thrun, R. Brooks, and H. Durrant-Whyte, Eds. San Francisco, CA: Springer, October 2005, pp. 214–234.

[27] ——, "Exactly sparse extended information filters for feature-based SLAM," *International Journal of Robotics Research*, vol. 26, no. 4, pp. 335–359, April 2007.