Visually Mapping the RMS Titanic: Conservative Covariance Estimates for SLAM Information Filters

Ryan M. Eustice*
Department of Naval Architecture
and Marine Engineering
University of Michigan
Ann Arbor, MI 48109 USA
eustice@umich.edu

Hanumant Singh
Department of Applied Ocean Physics
and Engineering
Woods Hole Oceanographic Institution
Woods Hole, MA 02543 USA
hanu@whoi.edu

John J. Leonard
Matthew R. Walter
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA, 02139 USA
{jleonard,mwalter}@mit.edu

Abstract

This paper describes a vision-based, large-area, simultaneous localization and mapping (SLAM) algorithm that respects the low-overlap imagery constraints typical of underwater vehicles while exploiting the inertial sensor information that is routinely available on such platforms. We present a novel strategy for efficiently accessing and maintaining consistent covariance bounds within a SLAM information filter, thereby greatly increasing the reliability of data association. The technique is based upon solving a sparse system of linear equations coupled with the application of constant-time Kalman updates. The method is shown to produce consistent covariance estimates suitable for robot planning and data association. Real-world results are reported for a vision-based, six degree of freedom SLAM implementation using data from a recent survey of the wreck of the RMS Titanic.

KEY WORDS—SLAM, data association, information filters, mobile robotics, computer vision, and underwater vehicles

*R. Eustice was with the Joint Program in Oceanographic Engineering of the Massachusetts Institute of Technology, Cambridge, MA, and the Woods Hole Oceanographic Institution, Woods Hole, MA during the tenure of this work.

The International Journal of Robotics Research
Vol. 25, No. 12, December 2006, pp. 1223-1242
DOI: 10.1177/0278364906072512

1. Introduction

This paper addresses the problem of precision navigation and mapping using low-overlap, high resolution image sequences obtained by unmanned underwater vehicles (UUVs). From a robotics science perspective, our primary contribution consists of an efficient algorithm for extracting consistent covariance bounds from SLAM information filters. From a robotics systems perspective, we demonstrate automatic visually augmented navigation (VAN) processing of a sequence of 866 images of the RMS Titanic, for a mission with a vehicle path length over 3 km long.

1.1. Motivation

A number of oceanographic applications share the requirement for high resolution imaging of sites extending over hundreds of meters. These include hydrothermal vent sites (German et al. 2004), cold seep sites (Hill et al. 2004), shipwrecks of archaeological significance (Ballard et al. 2002), coral reefs (Singh et al. 2004), and fisheries habitats (Reynolds et al. 2001). One of the significant challenges associated with such tasks is the requirement for precise and accurate navigation to ensure complete, repeatable coverage over the site of interest.

Traditionally, the oceanographic community has utilized three different methodologies (by themselves or in combina-

tion) to address navigation underwater: (1) transponder networks placed on the seafloor (Hunt et al. 1974), (2) ultra-short-baseline (USBL) range and bearing tracking systems (Milne 1983), and (3) ranging and inertial sensors on the underwater vehicle (Whitcomb et al. 1999a). Each of these methodologies trade off different aspects of accuracy, cost, and complexity.

Transponder networks provide bounded-error navigation on the seafloor, but come at the cost of the overhead required for the deployment and calibration of the individual transponders on the seafloor; these systems are also limited to providing updates every few seconds based upon the measured round-trip travel time between the vehicle and transponders.[1] Ship to vehicle USBL bearing measurements degrade as a function of water depth and are also limited to update rates that fall off with range. Inertial navigation systems, while providing consistent updates at a few hertz, yield unbounded errors as a function of distance traveled. For the most challenging tasks, the systems of choice are typically long-baseline transponder networks used in combination with inertial sensors on the underwater vehicle. Such systems ensure bounded-error surveys with rapid (a few hertz) update rates (Whitcomb et al. 1999b).

Our approach to autonomous, extended-duration, infrastructure-free navigation and mapping has been to explore a methodology that uses a vision-based SLAM approach, paralleling other state-of-the-art navigation research within the computer vision and robotics community (Davison 2003; Davison and Murray 2002; Roumeliotis 2002; Nister 2004; Se- et al. 2005; Repko and Pollefeeys 2005; van Gool et al. 2000; Pollefeys et al. 2004). While typical terrestrial structure-from-motion (SFM) approaches estimate both camera motion and 3D scene structure from a sequence of video frames, in our application the low degree of temporal image overlap (typically on the order of 35% or less) motivates us to focus on recovering pairwise measurements from spatially neighboring image frames. Hence, what differentiates our goal from the typical SFM approach is that we seek an algorithm that respects the constraints of low-overlap imaging, which is typical of underwater vehicles, while providing high precision, accurate, large-area navigation measurements when used in concert with onboard inertial measurements. In our approach (Eustice et al. 2004, 2006; Eustice 2005; Pizarro et al. 2003, 2004), pairwise registration of overlapping monocular imagery provides measurements of the 6-degree of freedom (DOF) relative coordinate transformation between poses modulo scale. These measurements are used as constraints in a recursive estimation framework that tries to determine the global poses consistent with the camera measurements and navigation prior. Our goal is the development of a real-time filtering algorithm, focused primarily on navigation instead of structure recovery, capable of scaling to large environments (image sequences consisting of thousands of key frames), while taking advantage of the complementary aspects of inertial sensing within the vision processing pipeline. We consider this problem from the information formulation of SLAM.

1.2. The Information Form

To our knowledge, the earliest related work that exploited the efficiency of the measurement update in the inverse covariance form was published by McLauchlan and Murray (1995), in the context of recursive SFM. This work was subsequently extended to realize a hybrid batch/recursive visual SLAM implementation that unified recursive SLAM and bundle adjustment (McLauchlan 2000). McLauchlan recognized the potential increase in efficiency that can be gained via approximations to maintain sparsity of the information matrix:

> It has long been known in the photogrammetry community, in the form of the equivalent normal formulation, that the [information] matrix ... takes a special sparse form in the context of reconstruction ... [However, in a recursive formulation] ... eliminating motion fills in the structure blocks. This has to be avoided to maintain update times proportional to n. So our *partial elimination adjustment* method is to ignore corrections that fill-in zero blocks, while applying the correction to the blocks which are already non-zero.

While the consistency implications of this approximation are unknown, in practice the method achieved results approaching those of a full batch solution for moderate duration image sequences.

Within the SLAM community, algorithms exploiting the sparse information representation for SLAM were first proposed by Thrun et al. (2002, 2003), Frese and Hirzinger (2001) and Frese (2004), with subsequent developments by Paskin (2002, 2003), Eustice et al. (2005a, 2006b), and Dellaert (2005). All of these methods exploit the observation that this representation is either sparse (Eustice et al. 2005a, 2006b; Dellaert 2005) or approximately sparse (Thrun et al. 2000; Frese and Hirzinger 2001; Frese 2004; Paskin 2002, 2003). The sparse representation allows for linear storage requirements and efficient fusion of sensor measurements. However, the recovery of covariances for data association and motion planning is a cubic operation if a naive approach is followed (i.e., matrix inversion).

The key issue on which we focus in this paper is the efficient recovery of *consistent* covariances from the information filter. Consistency and the coupled issue of computational efficiency are two of the key criteria that one would need to consider in developing a taxonomy of the many SLAM algorithms that have been proposed in recent years. While it is hard to define a single definition of consistency employed uniformly in the

1. The speed of sound in water is approximately 1500 mp/s.

prior literature on SLAM, intuitively, consistency reflects the goal that the error estimates computed by the filter should match the actual errors (Knight 2001).

In relation to SLAM, consistency of the error estimates is important for data association—determining the correspondences for measurements (Neira and Tardos 2001). This is important both in the context of local SLAM (e.g., detecting and tracking features), and in a global sense (e.g., closing loops). If the SLAM error estimates are too small (i.e., overconfident), then both of these tasks can become difficult. Before describing our approach for efficient recovery of consistent covariances bounds, we first review the basic characteristics of SLAM information filters.

2. SLAM Information Filters

A number of recent SLAM algorithms have explored reformulating the estimation problem within the context of an extended information filter (EIF), which is the dual of the extended Kalman filter (EKF) (Bar Shalom et al. 2001). The information form is often called the canonical or natural representation of the Gaussian distribution because it stems from expanding the quadratic in the exponential. The result is that rather than parametrizing the normal distribution in terms of its mean and covariance, $\mathcal{N}(\boldsymbol{\xi}_t; \boldsymbol{\mu}_t, \Sigma_t)$, it is instead parametrized in terms of its information vector and information matrix, $\mathcal{N}^{-1}(\boldsymbol{\xi}_t; \boldsymbol{\eta}_t, \Lambda_t)$, where

$$\Lambda_t = \Sigma_t^{-1} \quad \text{and} \quad \boldsymbol{\eta}_t = \Lambda_t \boldsymbol{\mu}_t. \tag{1}$$

2.1. Constant-time Measurement Updates

A well known and very attractive property of formulating SLAM in an EIF is that measurement updates are additive and efficient. This is in contrast to the quadratic complexity per update in the EKF. For example, assume the following general measurement function and its first-order linearized form:

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

where $\boldsymbol{\xi}_t \sim \mathcal{N}(\bar{\boldsymbol{\mu}}_t, \bar{\Sigma}_t) \equiv \mathcal{N}^{-1}(\bar{\boldsymbol{\eta}}_t, \bar{\Lambda}_t)$ is the time-propagated state vector, $\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathrm{R})$ is the white measurement noise, and H is the measurement Jacobian evaluated at the state mean, $\bar{\boldsymbol{\mu}}_t$. The EKF covariance update requires computing the Kalman gain and updating $\bar{\boldsymbol{\mu}}_t$ and $\bar{\Sigma}_t$ via (Bar Shalom et al. 2001):

$$\begin{aligned} \mathrm{K} &= \bar{\Sigma}_t \mathrm{H}^\top \left(\mathrm{H}\bar{\Sigma}_t \mathrm{H}^\top + \mathrm{R}\right)^{-1} \\ \boldsymbol{\mu}_t &= \bar{\boldsymbol{\mu}}_t + \mathrm{K}\left(\mathbf{z}_t - \mathbf{h}(\bar{\boldsymbol{\mu}}_t)\right) \\ \Sigma_t &= \left(\mathrm{I} - \mathrm{KH}\right)\bar{\Sigma}_t \left(\mathrm{I} - \mathrm{KH}\right)^\top + \mathrm{KRK}^\top. \end{aligned} \tag{2}$$

This calculation non-trivially modifies all elements in the covariance matrix resulting in quadratic computational complexity per update. In contrast, the corresponding EIF update is given by (Thrun et al. 2003):

$$\begin{aligned} \Lambda_t &= \bar{\Lambda}_t + \mathrm{H}^\top \mathrm{R}^{-1} \mathrm{H} \\ \boldsymbol{\eta}_t &= \bar{\boldsymbol{\eta}}_t + \mathrm{H}^\top \mathrm{R}^{-1}\left(\mathbf{z}_t - \mathbf{h}(\bar{\boldsymbol{\mu}}_t) + \mathrm{H}\bar{\boldsymbol{\mu}}_t\right). \end{aligned} \tag{3}$$

Equation (3) shows that the information matrix is additively updated by the outer product term $\mathrm{H}^\top \mathrm{R}^{-1} \mathrm{H}$. In general, this outer product modifies all elements of the predicted information matrix, $\bar{\Lambda}_t$, however a key observation is that the SLAM measurement Jacobian, H, is always sparse Thrun et al. 2003). For example, in our application we employ a view-based SLAM representation that uses monocular visual perception to extract relative-pose measurements from pairwise registration of overlapping images of the environment. Under this scenario, our state vector, $\boldsymbol{\xi}_t$, consists of a collection of historical poses sampled at image acquisition from our robot's trajectory:

$$\boldsymbol{\xi}_t = \begin{bmatrix} \mathbf{x}_1^\top & \dots & \mathbf{x}_i^\top & \dots & \mathbf{x}_n^\top & \mathbf{x}_r(t)^\top \end{bmatrix}^\top$$

where $\mathbf{x}_r(t)$ is the current robot state, $\mathbf{x}_i \equiv \mathbf{x}_r(t_i)$ for $t_i \leq t$ is a time-delayed trajectory sample, and n is the current number of views comprising our appearance-based map. Therefore, given a pair of images I_i and I_j, image registration provides a relative-pose measurement between states $\mathbf{x}_i$ and $\mathbf{x}_j$ resulting in a sparse Jacobian of the form:

$$\mathrm{H} = \begin{bmatrix} 0 \cdots & \frac{\partial \mathbf{h}}{\partial \mathbf{x}_i} & \cdots 0 \cdots & \frac{\partial \mathbf{h}}{\partial \mathbf{x}_j} & \cdots 0 \end{bmatrix}.$$

As a result, only the four-block elements corresponding to $\mathbf{x}_i$ and $\mathbf{x}_j$ of the information matrix need to be modified (i.e., $\bar{\Lambda}_{x_i x_i}$, $\bar{\Lambda}_{x_j x_j}$, and $\bar{\Lambda}_{x_i x_j} = \bar{\Lambda}_{x_j x_i}^\top$), due to the matrix outer product of equation (3). Since measurements only ever involve a fixed portion of the SLAM state vector, updates can be performed in constant-time.

2.2. Sparse Representation

Thrun et al. (2003) originally showed that the (filtered) feature-based SLAM information matrix empirically obeys a "close-to-sparse" structure when properly normalized. This observation spawned the development of a number of computationally efficient feature-based SLAM algorithms such as sparse extended information filters (SEIFs) (Thrun et al. 2003), thin junction-tree filters (TJTFs) (Paskin 2003), and Tree-Map filters (Frese 2004). These algorithms approximate the SLAM posterior by (effectively) eliminating small elements in the corresponding information matrix. The elimination of weak constraints results in a sparse representation allowing the development of efficient filtering algorithms that exploit the resulting sparse architecture. This empirical observation of weak inter-landmark constraints has recently been

given a solid theoretical foundation by Frese (2005) where he mathematically shows that inter-landmark information decays spatially at an exponential rate. This adds some justification for the sparseness approximations utilized in feature-based SLAM algorithms, though, recently Eustice et al. (2005b) have shown that sparsification can lead to global map inconsistency.

Alternatively to feature-based techniques, recent work by Eustice et al. (2005a, 2006b) show that for a view-based representation the SLAM information matrix is *exactly* sparse without any approximation (the root cause being the preservation of key historical samples from the robot's trajectory). The implication is that view-based SLAM systems can take advantage of the sparse information parametrization without incurring any sparse-approximation error. As an example, Figure 1 depicts the resulting information matrix associated with registering 866 images and fusing them with navigation data from a boustrophedon ROV survey of the RMS Titanic (Figure 2(a)). The off-diagonal elements in the information matrix correspond to cross-track camera measurements while the block-tridiagonal structure naturally arises from the first-order Markov vehicle process model. The wreck was surveyed amidships to stern and then amidships to bow (Figure 2(b)), which resulted in a large loop-closing event. This event is annotated in the information matrix of Figure 1 and appears as the far off-diagonal elements near the upper-right corner.

2.3. State Recovery

While the insight of sparsity has spawned the development of computationally efficient SLAM algorithms such as those mentioned, an issue countering the utility of the information filter is "how to gain efficient access to the state estimate and its uncertainty?" Referring back to (1) we see that the information parametrization embeds the state mean and covariance within the information vector and information matrix, respectively. State recovery implies that whenever we want to actually recover our state estimate for the purposes of motion planning, data association, map recovery, linearizing our process or observation models, etc., we must invert this relationship.

2.3.1. Recovering the Mean

Naïve recovery of our state estimate through matrix inversion results in cubic complexity and destroys any efficiency gained over the EKF. Fortunately, closer inspection reveals that recovery of the state mean, $\boldsymbol{\mu}_t$, can be posed more efficiently as solving the sparse, symmetric, positive-definite, linear system of equations:

$$\Lambda_t \boldsymbol{\mu}_t = \boldsymbol{\eta}_t. \tag{4}$$

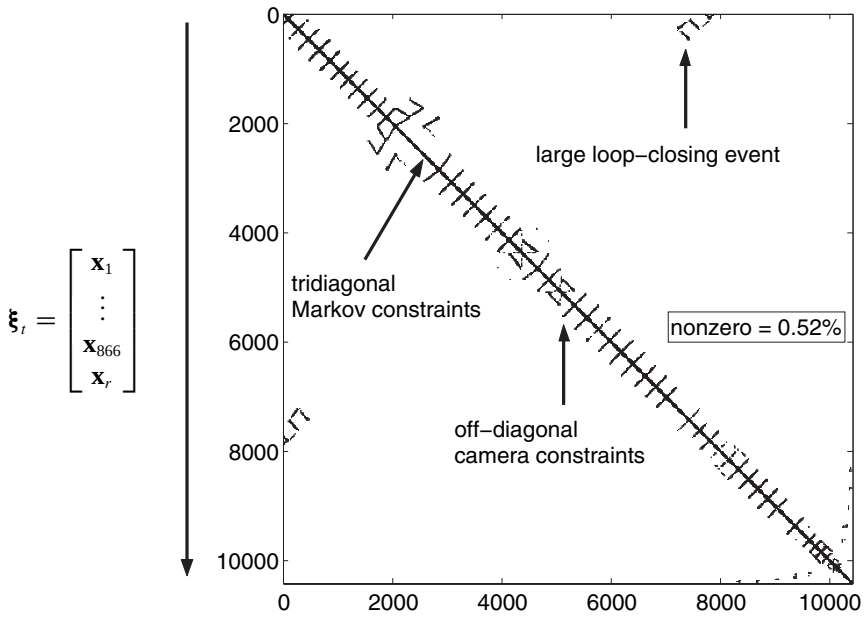
Such systems can be iteratively solved via the classic method of conjugate gradients (CG) (Shewchuk 1994). In general, CG can solve this system in n iterations (with $\mathcal{O}(n)$ cost per iteration) where n is the size of the state vector, and typically in many fewer iterations if the initialization is good (Konolige 2004). In addition, since the state mean, $\boldsymbol{\mu}_t$, often does not change significantly with each measurement update (excluding key events like loop-closure), this relaxation can take place over *multiple time steps* using a fixed number of iterations per update (Duckett et al. 2000; Thrun et al. 2003). Also, recently proposed multilevel relaxation SLAM algorithms, such as (Konolige 2004; Frese et al. 2005), claim linear asymptotic complexity. This computational reduction is achieved by sub-sampling poses and performing the relaxation over *multiple spatial scales*, which has the effect of improving convergence rates.

As an example of the state recovery efficiency that sparse linear systems can provide, Figure 3 details the amount of CPU-time utilized for full state recovery during the processing of the RMS Titanic dataset. The depicted results are for a batch recovery method whereby we used MATLAB's "left-divide" capability to solve (4) (i.e., $\boldsymbol{\mu}_t = \Lambda_t \backslash \boldsymbol{\eta}_t$) after the incorporation of each camera measurement. For comparison purposes, we fitted a least-squares power curve to the raw CPU-times and found that overall state recovery complexity scaled as $\mathcal{O}(n^{1.214})$ for this dataset, which, as Figure 3(b) shows, is similar to the $\mathcal{O}(n \log n)$ curve over several orders of magnitude in state size. Hence, a good implementation of any of the iterative relaxation techniques outlined in the preceding paragraph should perform even better.

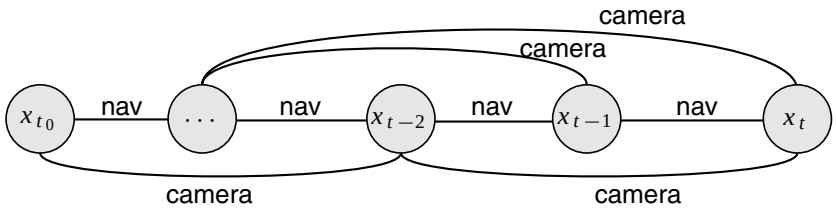
2.3.2. Recovering Covariance

The covariance matrix corresponds to the inverse of the information matrix, however, actually recovering the covariance via (1) is not practical since matrix inversion is a cubic operation. Additionally, while the information matrix can be a sparse representation for storage, in general, its inverse results in a *fully dense* covariance matrix despite any sparsity in the information form (Frese and Hirzinger 2001). This means that calculating the covariance matrix requires quadratic memory storage, which may become prohibitively large for very large maps (e.g., maps $\geq \mathcal{O}(10^5)$ elements). To illustrate this point, take for example the $10,404 \times 10,404$ information matrix shown in Figure 1, storing it in memory only requires 4.5 MB of double precision storage for the nonzero elements while its inverse requires over 865 MB.

Fortunately, recovering the entire covariance matrix is usually not necessary for SLAM as many of the data association and robotic planning decisions often do not require the full covariance matrix, but only the covariance over subsets of state variables (Dissanayake et al. 2001). Unfortunately, accessing only subsets of state variables in the information form is not an easy task. The covariance and information representations of the Gaussian distribution lead to very different computational characteristics with respect to the fundamental probabilistic



(a) The information matrix for the RMS Titanic survey.

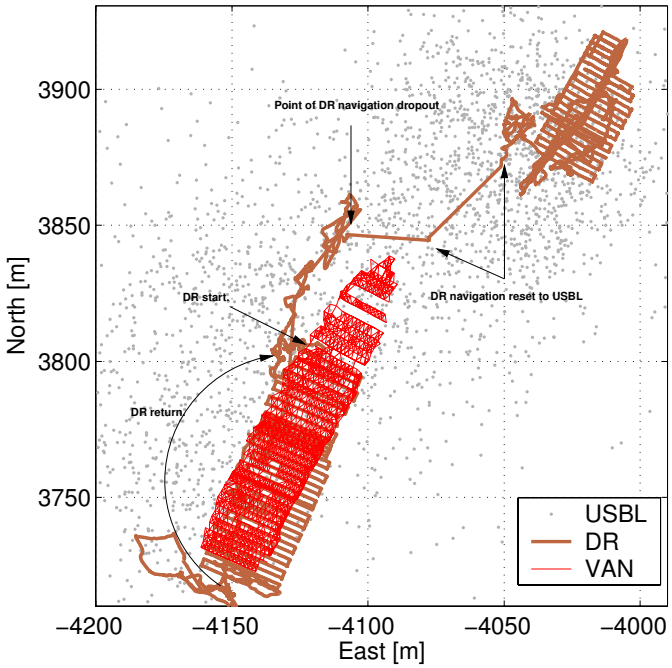


(b) A conceptual depiction of the Markov network.

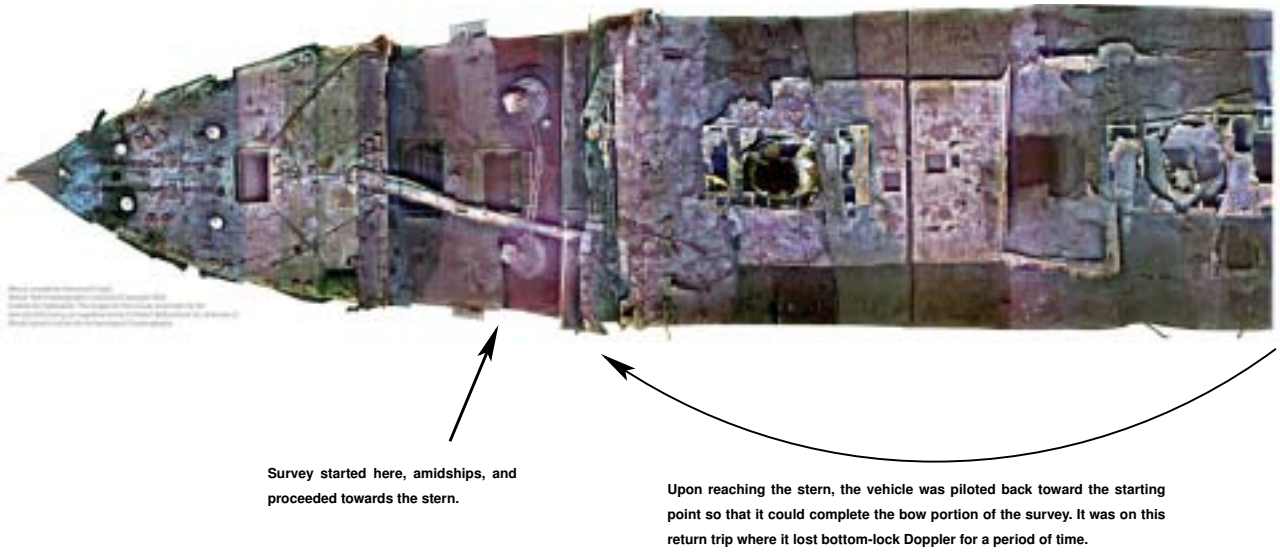
Fig. 1. View-based SLAM is exactly sparse. (a) This figure highlights the exact sparsity of the view-based SLAM information matrix using data from a recent ROV survey of the wreck of the RMS Titanic. In all there are 867 delayed-states where each state is a 12-vector consisting of 6-pose (i.e., Cartesian position and Euler attitude) and 6-kinematic components (i.e. linear and angular body-frame velocities). The resulting information matrix is a $10,404 \times 10,404$ matrix with only 0.52% nonzero elements. (b) The system diagram for a view-based representation. The model is comprised of a pose-graph where the nodes correspond to historical robot poses and edges represent either Markov (navigation) or non-Markov (camera) constraints.

operations of marginalization and conditioning (Table 1). For example, marginalization is easy in the covariance form, since it corresponds to extracting the appropriate sub-block from the covariance matrix, while in the information form it is hard, because it involves calculating the Schur complement over the variables we wish to keep (note that the opposite relation holds true for conditioning, which is easy in the information form and hard in the covariance form). Therefore, even though we may only need access to covariances over subsets of the state elements (Dissanayake et al. 2001) (and thus only have to invert a small information matrix related to the subset of variables we are interested in), accessing them in the information form requires marginalizing out most of the state vector resulting in cubic complexity due to matrix inversion in the Schur complement.

To sidestep this dilemma, Thrun et al. (2003) and Liu and Thrun (2003) proposed a data association strategy based upon using *conditional* covariances. Since conditional information matrices are easy to obtain in the information form (simply extract the sub-block over desired variables), their strategy is to choose an appropriate sub-block from the information matrix such that its inverse approximates the actual covariance for the subset of variables they are interested in. In particular, given two state variables of interest, $\mathbf{x}_i$ and $\mathbf{x}_j$, their approx-

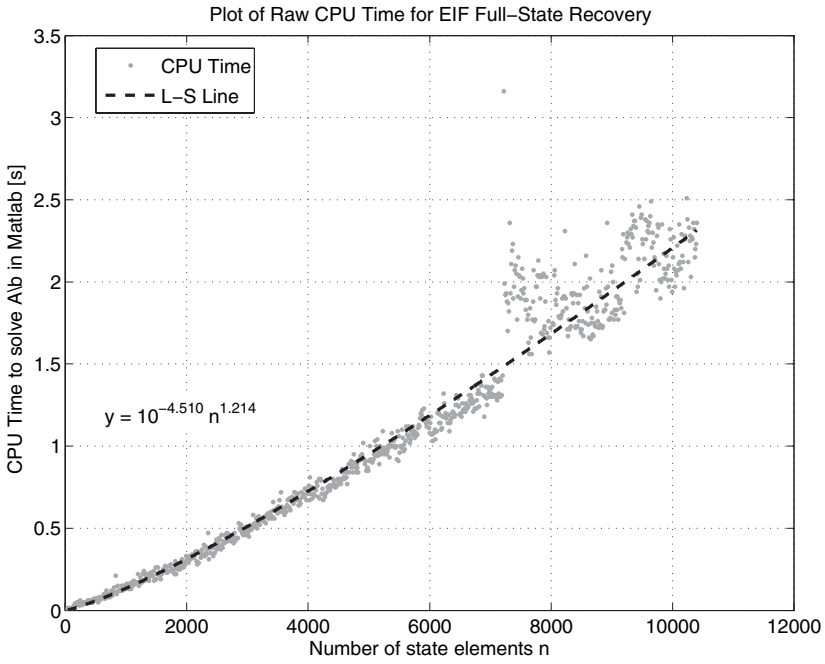


(a) Comparison of the different navigation results.

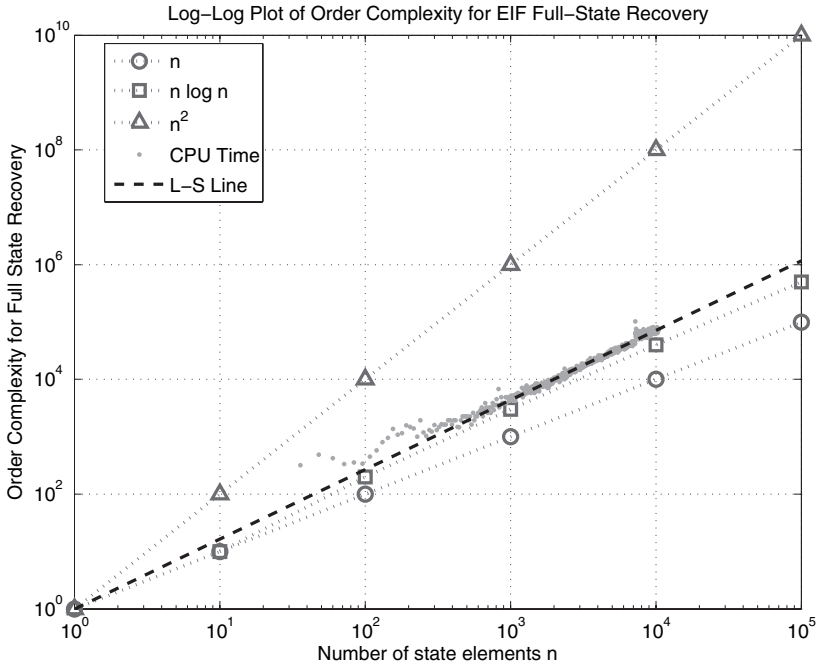


(b) Photomosaic of the RMS Titanic.

Fig. 2. Mapping results from a ROV survey of the RMS Titanic. (a) A XY plot comparing the raw dead-reckoned (DR) navigation data (brown), ship-board USBL tracking (gray), and visually reconstructed survey trajectory from a 6-DOF view-based SLAM information filter (red). The discontinuity in the DR trajectory is the result of navigation sensor dropouts. (b) A photomosaic of the RMS Titanic constructed from over 700 digital-still images. This photomosaic was generated independent of our algorithm and is presented for visualization purposes only as a representation of the data that serves as input. It is the result of semi-automatic processing with manual selection of a number of control points to guide the photomosaicking process and could be considered as a form of benchmark against which fully autonomous processing can be compared.



(a) Plot of CPU-time used during state recovery.



(b) Log-Log plot with benchmark curves overlaid for comparison.

Fig. 3. This figure portrays the CPU-time used in recovering the state mean estimate for the Titanic dataset shown as a function of the number of state elements, n. (a) This plot depicts the raw CPU-time (gray dots) for using a batch method to solve (4) after the incorporation of each camera measurement. Results are shown for MATLAB R13 on an Intel Pentium-4 3.4 GHz processor with 2048 MB of RAM. The batch method we refer to is MATLAB's "left-divide" capability, which employs a sparse Cholesky factorization followed by forward and backward substitution(Mathworks 2005). Overlaid in black is the least-squares fit power curve to the raw data showing that recovery complexity grows as $\mathcal{O}(n^{1.214})$ for this dataset. (b) A log-log graph of the measured complexity with benchmark n, $n \log n$, and n^2 curves overlaid for comparison. Note that over several orders of magnitude the raw CPU-time is similar to the $n \log n$ complexity curve.

Table 1. Summary of Marginalization and Conditioning Operations on a Gaussian Distribution Expressed in Covariance and Information Form

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_\alpha \\ \boldsymbol{\mu}_\beta \end{bmatrix}, \begin{bmatrix} \Sigma_{\alpha\alpha} & \Sigma_{\alpha\beta} \\ \Sigma_{\beta\alpha} & \Sigma_{\beta\beta} \end{bmatrix}\right) = \mathcal{N}^{-1}\left(\begin{bmatrix} \boldsymbol{\eta}_\alpha \\ \boldsymbol{\eta}_\beta \end{bmatrix}, \begin{bmatrix} \Lambda_{\alpha\alpha} & \Lambda_{\alpha\beta} \\ \Lambda_{\beta\alpha} & \Lambda_{\beta\beta} \end{bmatrix}\right)$$

	Marginalization	**Conditioning**
	$p(\boldsymbol{\alpha}) = \int p(\boldsymbol{\alpha}, \boldsymbol{\beta}) d\boldsymbol{\beta}$	$p(\boldsymbol{\alpha}\|\boldsymbol{\beta}) = p(\boldsymbol{\alpha}, \boldsymbol{\beta}) / p(\boldsymbol{\beta})$
Cov. Form	$\boldsymbol{\mu} = \boldsymbol{\mu}_\alpha$	$\boldsymbol{\mu}' = \boldsymbol{\mu}_\alpha + \Sigma_{\alpha\beta}\Sigma_{\beta\beta}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)$
	$\Sigma = \Sigma_{\alpha\alpha}$	$\Sigma' = \Sigma_{\alpha\alpha} - \Sigma_{\alpha\beta}\Sigma_{\beta\beta}^{-1}\Sigma_{\beta\alpha}$
Info. Form	$\boldsymbol{\eta} = \boldsymbol{\eta}_\alpha - \Lambda_{\alpha\beta}\Lambda_{\beta\beta}^{-1}\boldsymbol{\eta}_\beta$	$\boldsymbol{\eta}' = \boldsymbol{\eta}_\alpha - \Lambda_{\alpha\beta}\boldsymbol{\beta}$
	$\Lambda = \Lambda_{\alpha\alpha} - \Lambda_{\alpha\beta}\Lambda_{\beta\beta}^{-1}\Lambda_{\beta\alpha}$	$\Lambda' = \Lambda_{\alpha\alpha}$

imation selects the joint-Markov blanket $\mathbf{M}_i^+ \cup \mathbf{M}_j^+$ (i.e. $\mathbf{M}_k^+$ represents state variables *directly* connected to $\mathbf{x}_k$ in a graph theoretic sense within the information matrix) and additionally, if the intersection is null (i.e. $\mathbf{M}_i^+ \cap \mathbf{M}_j^+ = \emptyset$), variables along a path connecting $\mathbf{x}_i$ and $\mathbf{x}_j$ topologically. Their method then inverts this sub-block to obtain a covariance matrix for $\mathbf{x}_i$ and $\mathbf{x}_j$ conditioned on all other variables that have an indirect influence. They note that empirical testing indicates that their approximation seems to work well in practice for their application (Liu and Thrun 2003), despite the fact that using conditional covariances should result in an overconfident approximation.

3. Consistent Covariance Recovery

While recovering the mean estimate is a vital component for making real-world decisions when interacting with the environment, it alone is not always sufficient. For example, robotic tasks such a motion planning, data association, and loop-closing usually require some notion of the joint-uncertainty between the state estimates. Furthermore, estimates of how certain we are of map relations can have imperative implications on the action of the robot—quoting Uhlmann (1997):

> An autonomous vehicle controller, for example, might not take evasive action in response to an estimate that places the mean position of the vehicle at the edge of the road and an uncertainty of only one centimeter. But if the same estimate had an uncertainty of a meter, the controller would likely direct the vehicle toward the center of the lane to avoid the worst case possibility that it is actually off the road.

Our strategy for approximate covariance recovery from the information form is formulated upon gaining efficient access to meaningful values of covariance that are consistent with respect to the actual covariance obtained by matrix inversion. The motivation for a consistent approximation is that we guard against under-representing the uncertainty associated with our state estimates, which otherwise could lead to data association and robot planning errors. It is the access to meaningful values of joint-covariance for robot interaction, data association, and decision making in the information form that motivates our discussion. In this section we describe our strategy for obtaining covariance bounds within the context of our view-based SLAM application.

3.1. Efficiently Accessing The Robot's Covariance

We begin by noting that recovery of our state estimate, $\boldsymbol{\mu}_t$, from the information form already requires that we solve the sparse, symmetric, positive-definite system of equations (4) and moreover that this system can be solved in near linear time using the iterative techniques outlined in Section 2.3.1. Our covariance recovery strategy for the information form is based upon augmenting this linear system of equations so that the robot's covariance-column is accessible as well. Note that by definition $\Lambda_t \Sigma_t = \mathrm{I}$ and, therefore, by picking the i^{th} basis vector $\mathbf{e}_i$ from the identity matrix we can use it to selectively solve for a column of the covariance matrix, denoted Σ_{*i}, as

$$\Lambda_t \Sigma_t = \mathrm{I} \quad \Rightarrow \quad \Lambda_t \Sigma_{*i} = \mathbf{e}_i \quad \text{where} \quad \mathrm{I} = [\mathbf{e}_1, \ldots, \mathbf{e}_n].$$

To obtain the robot's covariance-column at any time step we simply augment our original linear system (4) to include an appropriate set of basis vectors, $\mathrm{E}_r = \{\mathbf{e}_r\}$, such that the solution to (5) provides access to the robot covariance-column, Σ_{*r}, along with the state mean:

$$\Lambda_t \begin{bmatrix} \boldsymbol{\mu}_t & \Sigma_{*r} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\eta}_t & \mathrm{E}_r \end{bmatrix}. \tag{5}$$

3.2. Consistent Covariances for Data Association

In this section we outline our strategy for recovering approximate joint-covariances useful for *data association*. Before we begin we want it to be clear to the reader that our technique for obtaining and maintaining these covariances should not be confused with the actual updating and mechanics of the information filter (Section 2). What we present in the following is a way of maintaining marginal *covariance bounds* (Figure 4) that are consistent with respect to the inverse of the information matrix. Furthermore, these covariances are used for data association *only* and are not in anyway involved in the actual information filter mechanics. With that being said we now present our algorithm.

3.2.1. Inserting a new map element

Given that (5) provides a mechanism for efficient access to the robot's covariance-column, Σ_{*r}, we exploit it to obtain useful covariance bounds for other map elements. For example, in VAN's view-based SLAM framework, whenever we insert a new image, I_i, into our view-based map, we correspondingly must augment our view-based SLAM state vector to include the new element, $\mathbf{x}_i$ (Eustice, Pizarro, and Singh 2004; Eustice, Singh, and Leonard 2005). This new state element, $\mathbf{x}_i$, corresponds to a sampling of our robot state at time t_i (i.e., $\mathbf{x}_i \equiv \mathbf{x}_r(t_i)$) and represents an estimate of where the robot was when it took that image. Since the two states are coincident at time t_i the covariance for $\mathbf{x}_i$ is

$$\Sigma_{ii} \equiv \Sigma_{rr},$$

which can be obtained by solving (5).[2] A well-known property of SLAM is that over time the covariance for $\mathbf{x}_i$ will *decrease* as new sensor measurements are incorporated and all map elements become fully correlated (Dissanayake et al. 2001). Therefore, storing $\tilde{\Sigma}_{ii} = \Sigma_{ii}$ as our initial covariance bound for $\mathbf{x}_i$ serves as a *conservative* approximation to the actual marginal covariance for all time, (i.e., $\tilde{\Sigma}_{ii} \geq \Sigma_{ii}(t)$).

3.2.2. Data association

In our application, the joint-covariance between the time-projected robot pose, $\mathbf{x}_r$, and any other map entry, $\mathbf{x}_i$, i.e.,

$$\bar{\Sigma}_{joint} = \begin{bmatrix} \bar{\Sigma}_{rr} & \bar{\Sigma}_{ir}^{\top} \\ \bar{\Sigma}_{ir} & \Sigma_{ii} \end{bmatrix},$$

2. Similarly, if instead we were using a feature-based SLAM approach, then the initial covariance bound for landmark, $\mathbf{L}_i$, could be computed as (Smith, Self, and Cheeseman 1990):

$$\Sigma_{ii} = \mathrm{G}_r \Sigma_{rr} \mathrm{G}_r^{\top} + \mathrm{G}_z \mathrm{R} \mathrm{G}_z^{\top},$$

where $\mathbf{g}(\mathbf{x}_r, \mathbf{z})$ is the feature initialization function, $\mathbf{z}$ and R are the measurement and its covariance, respectively, and G_r and G_z are the Jacobians.

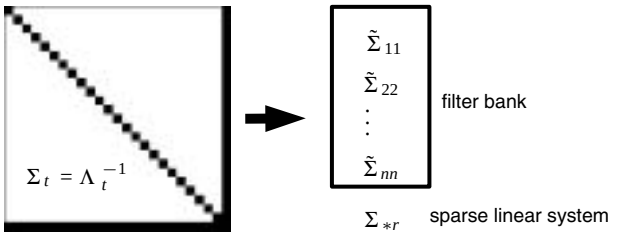


Fig. 4. The central idea behind our covariance recovery strategy is to maintain conservative estimates of map marginals using a parallel bank of filters, and augment that with knowledge of the robot's covariance-column obtained by solving a sparse linear system. This level of knowledge is sufficient to implement a standard maximum likelihood (i.e., nearest neighbor gating) data association strategy (Dissanayake et al. 2001). In the above figure, the matrix on the left depicts the elements of the covariance matrix that our algorithm attempts to recover; the rightmost column represents Σ_{*r} while columns 1 through n correspond to Σ_{*1} through Σ_{*n}, respectively.

is needed for two operations: link proposal and pose-constrained correspondence searches (Section 4.2). Link proposal (Eustice 2005; Eustice, Pizarro, and Singh 2006) corresponds to hypothesizing which images in our view-based map could potentially overlap with the current image being viewed by the robot, denoted I_r, and therefore could potentially be registered to generate a relative-pose measurement. The second operation, pose-constrained correspondence searches (Eustice, Pizarro, and Singh 2004, 2006), uses the relative-pose estimate between candidate images I_i and I_r to restrict the image-based correspondence search to probable regions based upon a two-view point transfer relation that exploits the epipolar constraint.

To obtain the actual joint-covariance, $\bar{\Sigma}_{joint}$, from the information matrix would require marginalizing out all other elements in the map except for $\mathbf{x}_r$ and $\mathbf{x}_i$, leading to cubic complexity in the number of eliminated variables. However, careful analysis shows that we can obtain a bounded approximation to $\bar{\Sigma}_{joint}$ at any time-step by using our solution from (5) to construct a conservative joint-covariance approximation, $\tilde{\bar{\Sigma}}_{joint}$, as

$$\tilde{\bar{\Sigma}}_{joint} = \begin{bmatrix} \bar{\Sigma}_{rr} & \bar{\Sigma}_{ir}^{\top} \\ \bar{\Sigma}_{ir} & \tilde{\Sigma}_{ii} \end{bmatrix}. \tag{6}$$

Here, $\bar{\Sigma}_{rr}$ and $\bar{\Sigma}_{ir}$ are exact and are extracted from $\bar{\Sigma}_{*r}$, while $\tilde{\Sigma}_{ii}$ is our stored conservative covariance bound. Note that (6) represents a valid positive-semidefinite and, therefore, consistent approximation satisfying:

$$\tilde{\bar{\Sigma}}_{joint} - \bar{\Sigma}_{joint} = \begin{bmatrix} 0 & 0 \\ 0 & \tilde{\Sigma}_{ii} - \Sigma_{ii} \end{bmatrix} \geq 0,$$

since $\tilde{\Sigma}_{ii} - \Sigma_{ii} \geq 0$. Given that (6) provides a consistent approximation to the true covariance, we can use it in our view-based VAN framework to compute conservative first-order probabilities of relative-pose in the usual way (i.e., $\mathbf{x}_{ri} = \ominus\mathbf{x}_r \oplus \mathbf{x}_i$ (Smith et al. 1990) for link hypothesis and correspondence searches.[3]

3.2.3. Updating our Covariance Bounds

Since $\tilde{\Sigma}_{ii}$ serves as a *conservative* approximation to the actual covariance, Σ_{ii}, for map element i, we would like to be able to place tighter bounds on it as we gather more measurement information. In fact, the careful reader will recognize that our SLAM information filter *is implicitly already doing this* for us. However, the issue is that extracting the actual filter bound, Σ_{ii}, from the information matrix is not particularly convenient. Note that while we could access Σ_{ii} by solving for the covariance-column Σ_{*i} using an appropriately chosen set of basis vectors, the reason for not doing this is that iteratively solving systems like (5) is efficient only when we have a good starting point (Duckett et al. 2000; Konolige 2004). In other words, when we solve (5) for the latest state and robot covariance-column, our previous estimates, $\boldsymbol{\mu}_t$ and Σ_{*r}, from the last time-step serve as good seed points and, therefore, typically only require a few iterations per time-step to update (excluding loop-closing events). In the case of solving for an arbitrary column, Σ_{*i}, we do not have a good *a priori* starting point and, therefore, convergence will be slower.

Our approach for tightening the bound, $\tilde{\Sigma}_{ii}$, is to use our joint-covariance approximation (6) and perform a simple constant-time Kalman update on a per re-observation basis (Algorithm 1). In other words, we only update the covariance bound, $\tilde{\Sigma}_{ii}$, when the robot re-observes $\mathbf{x}_i$ and successfully generates a relative-pose measurement, $\mathbf{z}_{ri}$, by registering images I_i and I_r. We then use that relative-pose measurement to perform a Kalman update (2) on the fixed-size state vector $\mathbf{y} = \left[\mathbf{x}_r^\top, \mathbf{x}_i^\top\right]^\top$ to obtain the new conservative bound, $\tilde{\Sigma}_{ii}^+$.[4]

Mathematically, the distribution over $\mathbf{y}$ corresponds to marginalizing out all elements in our state vector except for $\mathbf{x}_r$ and $\mathbf{x}_i$ as

$$p(\mathbf{y}) = \int_{\mathbf{x}_j \neq \{\mathbf{x}_r, \mathbf{x}_i\}} \mathcal{N}^{-1}\big(\bar{\boldsymbol{\eta}}_t, \bar{\Lambda}_t\big) d\mathbf{x}_j = \int_{\mathbf{x}_j \neq \{\mathbf{x}_r, \mathbf{x}_i\}} \mathcal{N}\big(\bar{\boldsymbol{\mu}}_t, \bar{\Sigma}_t\big) d\mathbf{x}_j. \quad (7)$$

3. Similarly, in a feature-based SLAM framework, knowledge of (6) is sufficient to implement the standard nearest-neighbor gating data association strategy (Dissanayake et al. 2001).

4. Similarly, in a feature-based SLAM framework we would use our sensor measurement, $\mathbf{z}_{ri}$, to perform an update on $\mathbf{y} = \left[\mathbf{x}_r^\top, \mathbf{L}_i^\top\right]^\top$ where $\mathbf{L}_i$ is landmark i.

Algorithm 1. Calculation of the marginal covariance bounds used for data association.

Require Σ_{*r} {initialize bound x}
  if $\mathbf{x}_i$ = new map element **then**
    store $\tilde{\Sigma}_{ii} \leftarrow \Sigma_{rr}$
  end if

Require $\bar{\boldsymbol{\mu}}_t$, $\bar{\Sigma}_{*r}$ {data association and bound update}
  for all $\mathbf{x}_i$ **do**
    $\tilde{\bar{\Sigma}}_{joint} \leftarrow \begin{bmatrix} \bar{\Sigma}_{rr} & \bar{\Sigma}_{ri} \\ \bar{\Sigma}_{ri} & \tilde{\Sigma}_{ii} \end{bmatrix}$
    compute link hypothesis
    if candidate link **then**
      do pose-constrained correspondence search on (I_i, I_r)
      if image registration success **then**
        do Kalman update on $\tilde{\bar{\Sigma}}_{joint}$ using measurement $\mathbf{z}_{ri}$
        store $\tilde{\Sigma}_{ii} \leftarrow \tilde{\Sigma}_{ii}^+$
      end if
    end if
  end for

The resulting distribution is then given by

$$p(\mathbf{y}) = \mathcal{N}\Big(\begin{bmatrix} \bar{\boldsymbol{\mu}}_r \\ \bar{\boldsymbol{\mu}}_i \end{bmatrix}, \begin{bmatrix} \bar{\Sigma}_{rr} & \bar{\Sigma}_{ir}^\top \\ \bar{\Sigma}_{ir} & \Sigma_{ii} \end{bmatrix}\Big). \quad (8)$$

Realizing that (6) already provides us with a consistent approximation to this distribution we have

$$\tilde{p}(\mathbf{y}) = \mathcal{N}\Big(\begin{bmatrix} \bar{\boldsymbol{\mu}}_r \\ \bar{\boldsymbol{\mu}}_i \end{bmatrix}, \begin{bmatrix} \bar{\Sigma}_{rr} & \bar{\Sigma}_{ir}^\top \\ \bar{\Sigma}_{ir} & \tilde{\Sigma}_{ii} \end{bmatrix}\Big), \quad (9)$$

where the only difference between the actual distribution, $p(\mathbf{y})$, and the approximation, $\tilde{p}(\mathbf{y})$, is the conservative marginal, $\tilde{\Sigma}_{ii}$. Using the measurement, $\mathbf{z}_{ri}$, we now perform a constant-time Kalman update (2) on (9) yielding the conditional distribution $\tilde{p}(\mathbf{y}|\mathbf{z}_{ri})$, from which we retain only the updated marginal bound, $\tilde{\Sigma}_{ii}^+$. This update is computed in constant-time for each re-observed map entry.

Note that by conceptually performing the marginalization step of (7) before computing the Kalman update, we have avoided any inconsistency issues associated with only storing the marginal bounds, $\tilde{\Sigma}_{ii}$, and not representing the intra-map correlations. This ensures that our update step will result in a consistent marginal bound for data association that will improve over time as we re-observe map elements.

4. Real-world Results

This section reports experimental results validating the consistency of our covariance recovery technique using field data collected from a remotely operated vehicle (ROV) survey of the RMS Titanic.

(a) ROV Hercules.

Measurement	Sensor	Precision
Roll/Pitch	Tilt Sensor	$\pm 0.1^\circ$
Heading	North-Seeking FOG	$\pm 0.1^\circ$
Body Frame Velocities	Acoustic Doppler	± 5 mm/s
Depth	Pressure Sensor	± 1 cm
Altitude	Acoustic Altimeter	± 10 cm
Down-looking Imagery	Calibrated 12-bit CCD	1 frame every 8 s

(b) Hercules' pose sensor characteristics.

Fig. 5. A depiction of the ROV Hercules and a table of its sensor characteristics (Coleman, Ballard, and Gregory 2003).

4.1. Experimental Setup

The wreck of the RMS Titanic was surveyed during the summer of 2004 by the deep-sea ROV Hercules (Coleman et al. 2003) (Figure 5) operated by the Institute for Exploration of the Mystic Aquarium. The ROV was equipped with a standard suite of oceanographic dead-reckon navigation sensors capable of measuring heading, attitude, altitude, XYZ bottom-referenced Doppler velocities, and a pressure sensor for depth. The vehicle surveyed the wreck athwartships while maintaining a constant altitude of approximately 7.5 m above the deck. The survey consisted of a boustrophedon trajectory at a horizontal speed of approximately 10 cm/s. The vehicle was equipped with a calibrated (intrinsic and extrinsic) stereo-rig consisting of two downward-looking 12-bit digital-still cameras that collected imagery at a rate of 1 frame every 8 seconds. This yielded an image sequence with slightly over 50% along-track (temporal) overlap and roughly 25% cross-track (spatial) overlap.

For the purposes of demonstration, the results reported herein were generated using imagery from the left stereo camera only (i.e., a monocular sequence). At no stage in the processing was the left-right stereo-pair information exploited. The purpose of this self-imposed restriction to a monocular image sequence is to demonstrate the general applicability of our VAN methodology.

Figure 2 summarizes the survey pattern and compares the different navigation methods used for localizing the vehicle. For real-time control, the vehicle integrated bottom-lock Doppler velocity measurements to obtain a dead-reckoned estimate of XY position. Additionally, ship-based USBL tracking provided range and bearing fixes to the vehicle used for shipboard tracking of the ROV. Since the wreck lies at a depth of approximately 3750 m, the large ship-to-vehicle moment arm, coupled with angular error in the USBL bearing measurements, resulted in an almost useless measurement of vehicle tracking as indicated by the widely distributed scatter of fixes in Figure 2(a).

The survey pattern consisted of a boustrophedon trajectory containing both temporal (along-track) and side-to-side (cross-track) overlap. The survey started amidships and proceeded towards the stern, as depicted in both Figure 2(a) and (b). Upon reaching the aft portion of the wreck, the camera was turned off and the vehicle was piloted back towards the starting point. During its return trip, the vehicle lost bottom-lock Doppler velocity measurements for a period of time, and therefore, was unable to dead-reckon integrate its vehicle position during this time period—this is the cause of the discontinuity denoted in the brown trajectory of Figure 2(a). After the vehicle returned near its starting point, the camera was turned back on and the vehicle completed the survey by mapping the bow of the wreck.

4.2. Image Processing

Our pairwise image registration algorithm assumes an ideal (i.e. distortion compensated) calibrated camera with known extrinsic reference frames (e.g., vehicle to camera). The view-based SLAM result depicted in Figure 2(a) (red) is the result of fusing onboard sensor-derived measurements with pairwise camera-derived relative-pose constraints (i.e., USBL measurements were not used in the SLAM estimate). These camera constraints were generated using a state-of-the-art feature-based image registration approach (Hartley and Zisserman 2000) founded upon:

1. Extract a combination of both Harris (Harris and Stephens 1988) and SIFT (Lowe 2004) interest points from each image. For the Harris points, we exploit our navigation prior to apply an orientation normalization to the interest regions by warping via the infinite homography (Hartley and Zisserman 2000), and then compactly encode using Zernike moments (Pizarro 2004).

2. Propose candidate image pairs based upon a probable measure of spatial proximity (Eustice 2005; Eustice et al. 2006a). Our link hypothesis strategy computes a probability of image overlap using our current pose-graph estimate and measured scene altitude. Under this scheme, we set thresholds for minimum and maximum percentage image overlap and then compute a first-order probability associated with whether or not the distance between the camera pair falls within these constraints. This calculation serves as the basis of our automatic link hypothesis strategy, where all frames in our view-based map are checked to see whether or not they could overlap with the current robot view (i.e., linear complexity in the number of views). The k most likely candidates ($k = 5$ in our application) are then sent to our image registration module for comparison.

3. Establish putative correspondences between overlapping candidate image pairs based upon similarity and a pose-constrained correspondence search (Eustice et al. 2004). We use the epipolar geometry constraint expressed as a two-view point transfer model to restrict the correspondence search to probable regions. These regions are determined by our pose prior and altitude, and are used to confine the interest point matching to a small subset of candidate correspondences. The benefit of this approach is that it simultaneously relaxes the demands of the feature descriptor while at the same time improves the robustness of similarity matching.

4. Employ a statistically robust least median of squares (LMedS) (Rousseeuw and Leroy 1987) registration methodology with regularized sampling (Zhang 1998) to extract a consistent inlier correspondence set. For this task we use a 6-point Essential matrix algorithm (Pizarro et al. 2003) as the motion-model constraint.

5. Solve for a relative-pose estimate using the inlier correspondence set and Horn's relative orientation algorithm (Horn 1990) initialized with samples from our orientation prior.

6. Carry out a two-view maximum likelihood estimate (MLE) refinement to extract the optimal 5-DOF relative-pose constraint (i.e., azimuth, elevation, Euler roll, Euler pitch, Euler yaw) and first-order parameter covariance based upon minimizing the reprojection error over all inliers (Hartley and Zisserman 2000).

Figure 6 depicts a block-diagram of the overall pairwise image processing pipeline. For further details on our systems-level image processing, including link hypothesis and two-view pose-constrained correspondence searches, the reader is referred to Eustice et al. (2004, 2006a), Eustice (2005), Pizarro et al. (2003, 2004).

4.3. Experimental Results

In Figure 7(a) we see a 2D view of the final pose-constraint network with Figure 7(b) providing a zoomed view of the boxed region. This inset facilitates comparison of the marginal covariance bounds estimated by our algorithm to the actual bounds obtained by matrix inversion. Note that all estimated bounds were verified to indeed be consistent with the actual bounds obtained by matrix inversion. This was done by performing Cholesky decomposition on their difference to establish positive definiteness. Because our algorithm only updates the bounds on a per re-observation basis, some of the estimated covariance bounds (gray) are tighter approximations than others to the actual filter bounds (green). This characteristic is a result of whether or not the robot is sufficiently well-localized when it re-observes an image, if it is, then the covariance bound for the re-observed map element improves.

Moving on, Figure 8 provides a quantitative assessment comparing the covariance bounds obtained by our algorithm

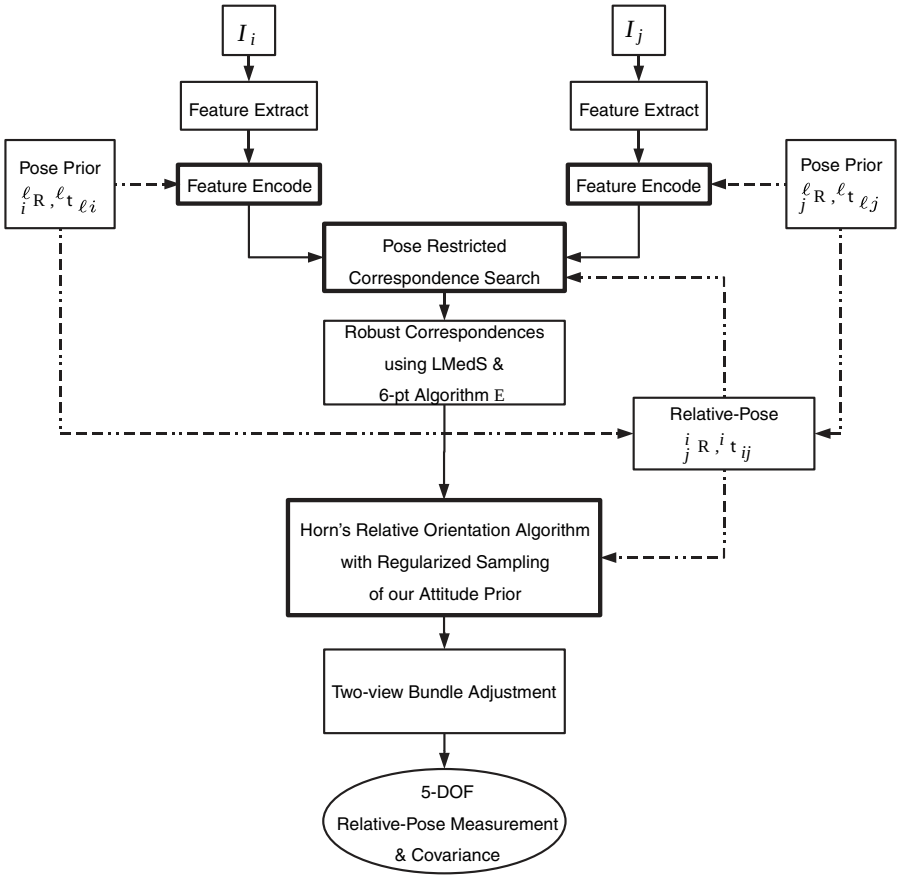
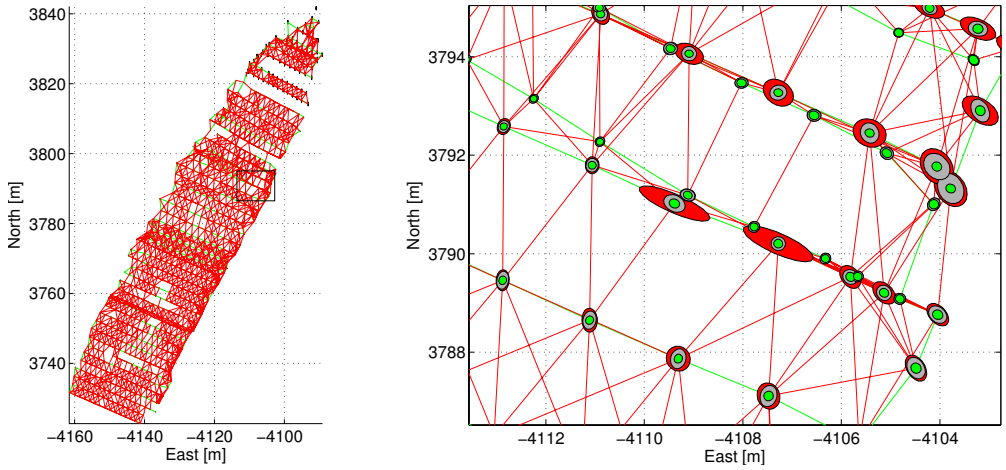


Fig. 6. An overview of the pairwise image registration engine. Dashed lines represent additional information provided by our SLAM state estimate, while bold boxes represent our systems-level extensions to a typical feature-based registration framework. Given two images I_i and I_j, we detect features using a combination of Harris and SIFT interest operators. For the Harris points, we exploit our navigation prior to orientation normalize the interest regions by warping via the infinite homography, H_{∞}. For each feature, we establish a putative match based upon similarity and a novel pose-constrained correspondence search. A 6-point essential matrix algorithm employed within a statistically robust LMedS strategy extracts an inlier correspondence set. Using this set we initialize our relative-pose estimate using Horn's relative orientation algorithm with regularized sampling from our orientation prior. We then refine this estimate in a two-view bundle adjustment step based upon minimizing the reprojection error over all inliers.

to the bounds obtained by inverting only the Markov Blanket as proposed in Liu and Thrun (2003) and Thrun et al. (2003). To provide a fair assessment, we choose to evaluate the *relative* uncertainty between the robot, $\mathbf{x}_r$, and any other map element, $\mathbf{x}_i$. Our justification for this metric is that selecting only the Markov Blanket results in a conditional covariance that does not accurately reflect *global* map uncertainty, but rather *relative* map uncertainty. Using the information matrix of Figure 1, for each map element, $\mathbf{x}_i$, we computed the first-order relative-pose (i.e, $\mathbf{x}_{ri} = \ominus\mathbf{x}_r \oplus \mathbf{x}_i$) and associated covariance matrix between it and the robot. For our metric we chose to compute the ratio between the determinant of the approximated covariance, to the determinant of the actual covariance (obtained by matrix inversion), and then take the logarithm:

$$\varepsilon_i = \log \frac{|\tilde{\Sigma}_{ii}|}{|\Sigma_{ii}|}.$$

With this metric, values greater than zero are conservative, values less than zero are overconfident, and zero indicates ideal. Figure 8(a) plots this metric evaluated for the the Titanic dataset. It shows that our method yields a conservative covariance approximation while the Markov blanket produces an overconfident estimate. Figure 8(b) shows the same result, but in histogram form. Note that the histogram for our approximation tends to be more centered near zero (i.e., closer to ideal) than the Markov blanket approximation (Figure 8(b)).



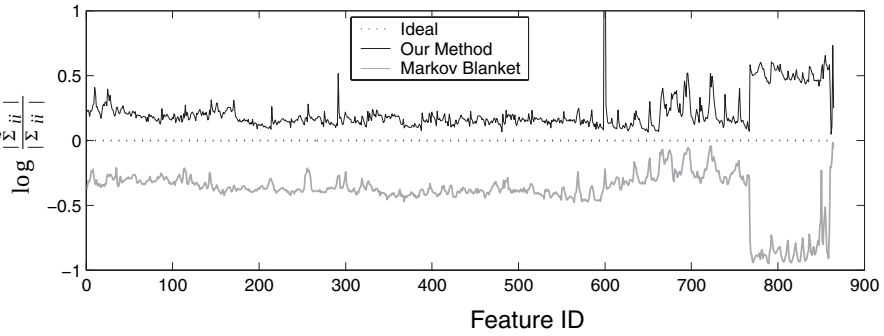
(a) 2D view of the final pose network.

(b) Zoomed view of the boxed region.

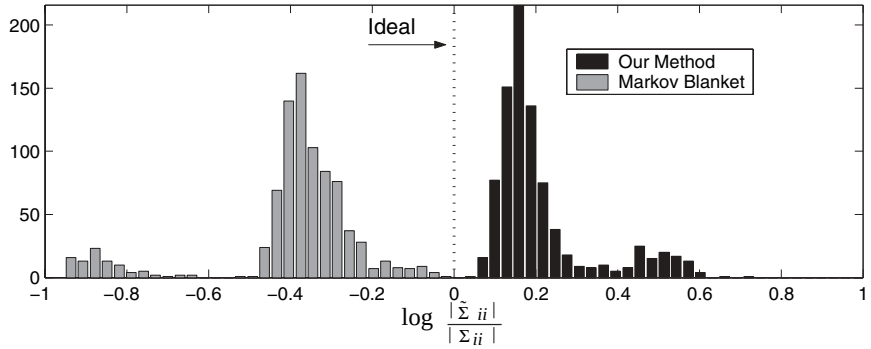
Fig. 7. Comparison of the proposed covariance recovery technique to that obtained by matrix inversion. (a) A top-down view of the final XY pose-constraint network associated with using 866 images to provide 3494 camera constraints; 3σ bounds are shown. Green links represent temporally consecutive registered image pairs while red links represent spatially registered image pairs. (b) A zoomed view illustrating the consistency of the data association bounds generated by our algorithm. Note that for this plot the 3σ bounds have been inflated by a factor of 30 for visualization. In this plot we have: (1) the initial covariance bounds associated at pose insertion (red), (2) the marginal covariance bounds based upon constant-time Kalman updates (gray), and (3) the actual marginal covariance bounds obtained by matrix inversion (green). Note that all actual filter bounds (green) lie within the estimated bounds (gray); this claim was verified for the entire dataset using Cholesky decomposition on their difference.

Next, Figure 9 demonstrates the actual value of this conservative approximation within the context of pose-constrained image registration. In particular, two candidate pairs of underwater images are shown with their predicted epipolar geometry (instantiated from the state estimate). Recall that for a calibrated camera, the epipolar geometry is specified by the relative camera pose and defines a 1D search constraint (Hartley and Zisserman 2000). However, when the relative-pose is uncertain, this 1D search constraint becomes a search *region* (Eustice et al. 2004, 2006a; Eustice 2005). Figures 9(a)–(c) depict a case where the Markov blanket approximation fails due to its overconfident covariance estimate. This failure is indicated by the fact that its 6σ confidence search region does not contain the true image correspondence while, in contrast, the regions computed using both the actual covariance and our conservative approximation do. Meanwhile, Figures 9(d)–(f) highlight that the amount of overconfidence in the Markov blanket approximation is unpredictable, since for a different image pair it produces comparable results to the other methods. This implies that we cannot simply inflate the Markov covariance estimate to compensate for its overconfidence, which furthermore stresses the importance of our conservative approximation algorithm.

Finally, while there is no ground-truth for this dataset, we can to some extent corroborate the accuracy of the recovered global poses by pairwise triangulating scene structure using the pairwise image correspondences and the VAN estimated vehicle poses. Figure 10 displays this result along with a Delaunay triangulated surface fitted to the reconstructed point cloud. The result is a coarse 3D surface model of the Titanic wreck as she now sits at the bottom of the ocean. Using this model we can construct a true overhead photomosaic of the wreck by back-projecting the imagery onto the 3D surface map as shown in Figure 10(c). For this rendering, images are simply draped over the mesh without any blending. This facilitates visual inspection of the quality of the reconstruction, since objects extending over multiple image seams should appear registered. This *quantitative*, *fully-automatic* result represents a significant advancement over the *qualitative*, *semi-automatic* mosaic presented in Figure 2(b). Moreover, it suggests that the VAN framework may have applicability in the SFM community, where it could be used to provide a consistent global pose estimate and coarse surface reconstruction useful for seeding an optimal offline bundle adjustment and dense surface reconstruction.



(a) Plot comparison of the different covariance approximation magnitudes.



(b) Histogram comparison of the different covariance approximation magnitudes.

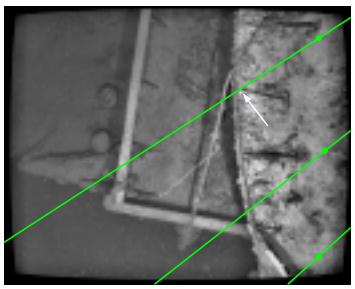
Fig. 8. A quantitative comparison of the different covariance recovery techniques using the information matrix of Figure 1. These plots compare the Markov blanket covariance approximation to the results of our method, both of which are shown relative to the actual covariance obtained by matrix inversion. For each method and state entry $\mathbf{x}_i$, we compute its relative-pose to the robot, $\mathbf{x}_r$, (i.e., $\mathbf{x}_{ri} = \ominus\mathbf{x}_r \oplus \mathbf{x}_i$) and associated first-order covariance. We then plot the log of the ratio of the determinant of the approximated covariance to the determinant of the actual covariance to facilitate comparison; conservative approximations take on positive values while overconfident approximations take on negative values. (a) Plot of the log ratio verses feature id for all $\mathbf{x}_i$. Note that a value of zero is ideal as this would indicate a ratio of one. (b) Same data as above but presented in histogram form. Both plots cross-validate that the method reported in this paper is conservative while the Markov blanket method is overconfident.

5. Conclusions and Suggestions for Future Work

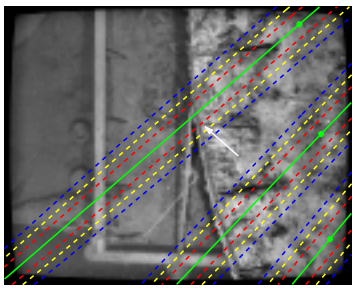
This article reported a novel algorithm for efficiently extracting conservative covariance bounds from SLAM information filters. Our results were presented within the context of an actual robotic mapping survey of the RMS Titanic. In all we visually mapped a region covering 3100^2 m (convex hull) with a traversed 3D path length over 3.4 km. This achievement represents a significant step forward in employing vision-based SLAM techniques to real-world mapping contexts.

We demonstrated that our covariance recovery method produces a conservative covariance approximation, with respect to the actual covariance obtained by inverting the information matrix, for the joint robot/map marginals. This covariance approximation is useful for real-world tasks such as nearest neighbor data association, image link hypothesis, and pose-constrained image registration. The method's complexity scales asymptotically linear with map size as measured by solving for the robot's covariance-column coupled with constant-time Kalman updates for re-observed map elements.

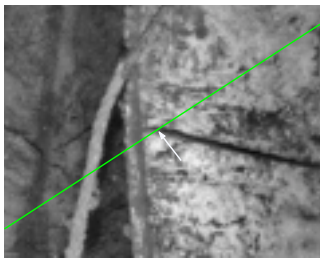
While this new technique addresses the lacuna regarding data association in SLAM information filters, it does not, however, categorically solve it. The current method efficiently extracts the joint-covariance marginals only between the robot and individual map-elements. Although this provides enough knowledge to implement a standard nearest-neighbor gating data association strategy (Newman 1999; Dissanayake et al. 2001), it is too impoverished of a representation for more sophisticated data association strategies that require knowledge of inter-landmark covariances, such as joint compatibility

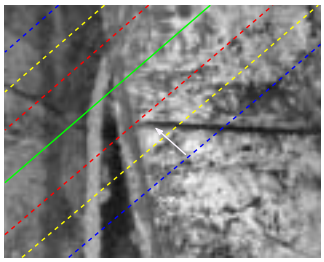


(a) Query image and its epipolar geometry.

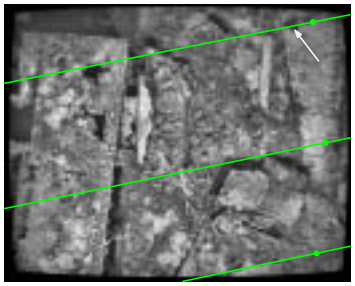


(b) Candidate image and its search regions.

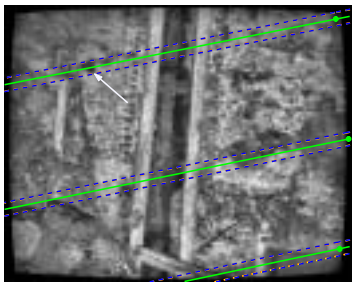




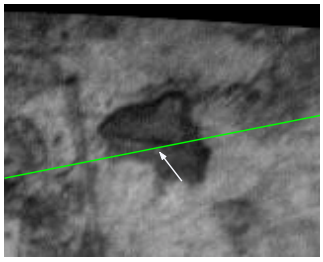
(c) Zoomed view. For this case, the Markov blanket approximation fails.

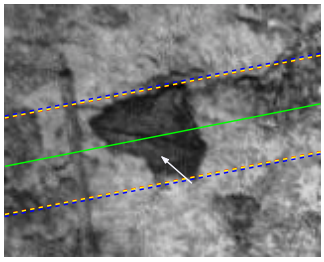


(d) Query image and its epipolar geometry.



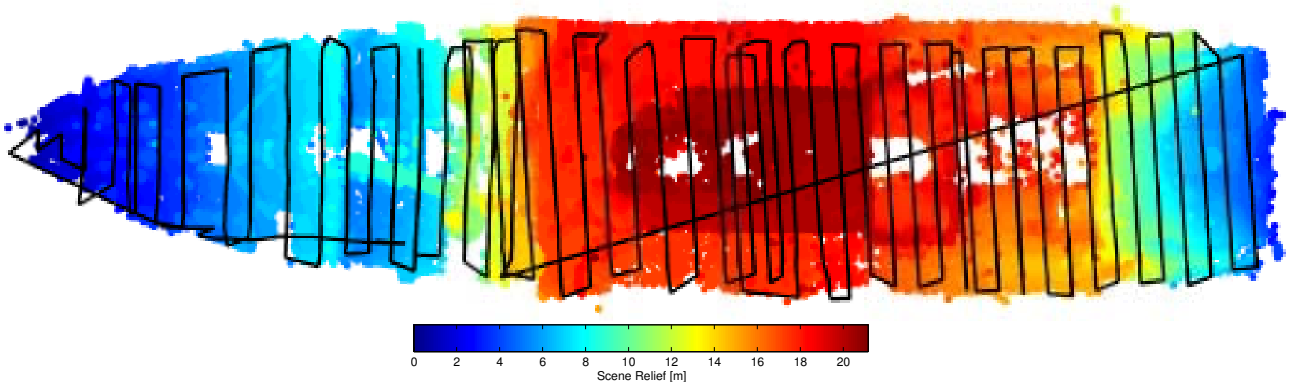
(e) Candidate image and its search regions.



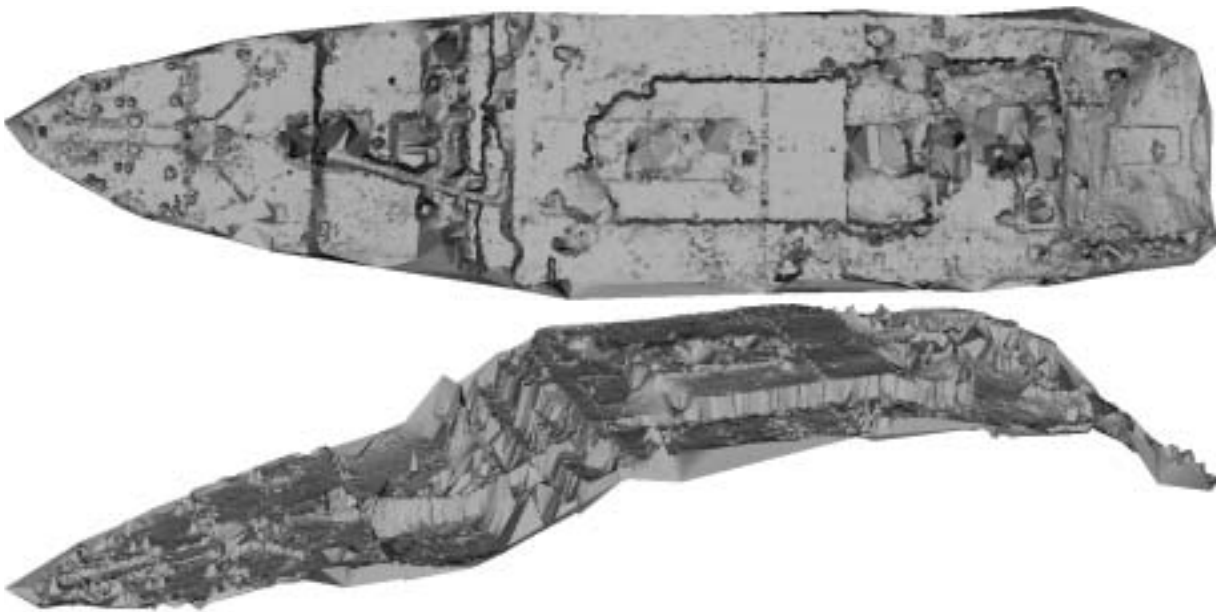


(f) Zoomed view. For this case, all three methods are comparable.

Fig. 9. Performance of the different covariance recovery techniques within the context of image registration. (a)–(b) These images are a proposed candidate pair for image registration. Image (a) represents the query image as viewed by the robot, and overlaid on top is the predicted epipolar geometry (green) instantiated from our state estimate. Image (b) is the proposed candidate for image registration, and overlaid on top are the pose-constrained correspondence search regions for 6σ confidence bounds. The different colored regions correspond to the three covariance recovery methods presented in this paper: (1) our conservative method (blue), (2) the actual covariance obtained by inverting the information matrix (yellow), and (3) the Markov blanket recovery technique (red). (c) These images show a zoomed view of the true correspondence, which is indicated by the white arrow. Careful inspection reveals that the Markov blanket search region (red) does not contain the true correspondence. In contrast, both the actual covariance (yellow) and our covariance approximation (blue) do. (d)–(f) These figures depict the same demonstration as (a)–(c), but for a different image pair. In this example, all three methods produce comparable results. This highlights the unpredictable nature of the Markov blanket approximation.



(a) Overhead view of the triangulated point cloud and tracklines (approximately 133 m $\times$ 30 m).



(b) Overhead and oblique view of the Delaunay triangulated surface.

(c) Overhead view of the texture-mapped surface.

Fig. 10. The triangulated point cloud, resulting Delaunay surface, and texture-mapped rendering for the RMS Titanic. (a) The triangulated point cloud calculated using VAN pose estimates and pairwise correspondences. Overlaid in black are the tracklines connecting sequential trajectory poses. (b) The resulting Delaunay triangulated surface. (c) The textured-mapped surface as computed by back-projecting the images onto the Delaunay mesh (the tiling artifact is due to the overlay of images without blending).

branch and bound (JCBB) (Neira and Tardos 2001). A second limitation of the current method is that covariance bounds are only updated upon a re-observation basis; in other words, map correlation is not exploited to improve the marginal bounds for unobserved map elements that share correlation with the current observation. This may lead to overly conservative covariance bounds for map elements that have not been directly observed for a long period of time.

Acknowledgments

We are grateful to Dr. Robert D. Ballard for providing us with the RMS Titanic dataset and wish to thank Professor Seth Teller for his discussions regarding large-area, scalable SLAM. This work was funded in part by the Center for Subsurface Sensing and Imaging Systems (CenSSIS) Engineering Research Center of the National Science Foundation under grant EEC-9986821, in part by the Woods Hole Oceanographic Institution through a grant from the Penzance Foundation, and in part by a National Defense Science and Engineering Graduate (NDSEG) Fellowship awarded through the Department of Defense.

References

Ballard, R., Stager, L., Master, D., Yoerger, D., Mindell, D., Whitcomb, L., Singh, H., and Piechota, D. 2002. Iron age shipwrecks in deep water off Ashkelon, Israel. *American Journal Archaeology* 106(2):151–168.

Bar-Shalom, Y., Rong Li, X., and Kirubarajan, T. 2001. *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, Inc., New York.

Coleman, D., Ballard, R., and Gregory, T. 2003. Marine archaeological exploration of the Black Sea. *Proceedings of IEEE/MTS OCEANS Conference and Exhibition*, Vol. 3, September, pp. 1287–1291.

Davison, A. 2003. Real-time simultaneous localisation and mapping with a single camera. *Proceedings of IEEE International Conference Computer Vision*, pp. 1403–1410.

Davison, A. and Murray, D. 2001. Simultaneous localisation and map-building using active vision. *IEEE Transactions Pattern Analysis and Machine Intelligence* 24(7):865–880.

Dellaert, F. 2005. Square root SAM. *Proceedings of Robotics: Science & Systems*. MIT Press, Cambridge, MA, pp. 177–184.

Dissanayake, M., Newman, P., Clark, S., Durrant-Whyte, H., and Csorba, M. 2001. A solution to the simultaneous localization and map building (SLAM) problem. *IEEE Transactions Robotics Automation* 17(3):229–241.

Duckett, T., Marsland, S., and Shapiro, J. 2000. Learning globally consistent maps by relaxation. *Proceedings of IEEE International Conference Robotics Automation*, San Francisco, CA, April, pp. 3841–3846.

Eustice, R., Pizarro, O., and Singh, H. 2004. Visually augmented navigation in an unstructured environment using a delayed state history. *Proceedings of IEEE International Conference Robotics Automation*, Vol. 1, New Orleans, USA, April, pp. 25–32.

Eustice, R. 2005. Large-area visually augmented navigation for autonomous underwater vehicles. PhD dissertation, Massachusetts Institute of Technology/Woods Hole Oceaonographic Institution Joint Program, June.

Eustice, R., Pizarro, O., and Singh, H. 2006a. Visually augmented navigation for autonomous underwater vehicles. *IEEE Journal Oceanic Engineering*, 2006, Accepted, To Appear.

Eustice, R., Singh, H., and Leonard, J. 2005a. Exactly sparse delayed-state filters. *Proceedings of IEEE International Conference Robotics Automation*, Barcelona, Spain, pp. 2428–2435.

Eustice, R., Singh, H., and Leonard, J. 2006b. Exactly sparse delayed-state filters for view-based SLAM. *IEEE Transactions Robotics*, Accepted, To Appear.

Eustice, R., Walter, M., and Leonard, J. 2005b. Sparse extended information filters: insights into sparsification. *Proceedings of IEEE/RSJ International Conference Intelligent Robot Systems*, pp. 641–648.

Frese, U. and Hirzinger, G. 2001. Simultaneous localization and mapping—a discussion. *Proceedings of IJCAI Workshop: Reasoning with Uncertainty in Robotics*, Seattle, WA, pp. 17–26.

Frese, U. 2004. Treemap: an O(Log N) algorithm for simultaneous localization and mapping. *Spatial Cognition IV*, C. Freksa, Ed., Springer Verlag.

Frese, U. 2005. A proof for the approximate sparsity of SLAM information matrices. *Proceedings of IEEE International Conference Robotics Automation*, Barcelona, Spain, pp. 331–337.

Frese, U., Larsson, P., and Duckett, T. 2005. A multilevel relaxation algorithm for simultaneous localisation and mapping. *IEEE Transactions Robotics* 21(2): 1–12.

German, C., Connelly, D., Prien, R., Yoerger, D., Jakuba, M., Bradley, A., Shank, T., Edmonds, H., and Langmuir, C. 2004. New techniques for hydrothermal exploration: in situ chemical sensors on AUVs—preliminary results from the Lau Basin. *EOS: Transactions American Geophysical Union Fall Meeting Supplement*, December.

Harris, C. and Stephens, M. 1998. A combined corner and edge detector. *Proceedings of 4th Alvey Vision Conference*, Manchester, UK, pp. 147–151.

Hartley, R. and Zisserman, A. 2000. *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Hill, J., Driscoll, N., Weissel, J., Kastner, M., Singh, H., Cormier, M., Camilli, R., Eustice, R., Lipscomb, R., McPhee, N., Newman, K., Robertson, G., Solomon, E., and Tomanka, K. 2004. A detailed near-bottom survey of large gas blowout structures along the US Atlantic

shelf break using the autonomous underwater vehicle (AUV) SeaBED. *EOS: Transactions American Geophysical Union Fall Meeting Supplement*.

Horn, B. 1990. Relative orientation. *International Journal Computer Vision* 4(1):59–78.

Hunt, M., Marquet, W., Moller, D., Peal, K., Smith, W., and Spindel, R. 1974. An acoustic navigation system. Woods Hole Oceanographic Institution, Technical Report WHOI-74-6, December.

Knight, J. 2001. Computationally tractable SLAM. University of Oxford, Technical Report OUEL 2232/2001.

Konolige, K. 2004. Large-scale map-making. *Proceedings of AAAI National Conference Artificial Intelligence*, San Jose, CA, pp. 457–463.

Liu, Y. and Thrun, S. 2003. Results for outdoor-SLAM using sparse extended information filters. *Proceedings of IEEE International Conference Robotics Automation*, Vol. 1, September, pp. 1227–1233.

Lowe, D. 2004. Distinctive image features from scale-invariant keypoints. *International Journal Computer Vision* 60(2):91–110.

The Mathworks. 2005. MATLAB function reference: mldivide, www.mathworks.com/access/helpdesk/help/techdoc/ref/mldivide.html, Aug. 2005.

McLauchlan, P. and Murray, D. 1995. A unifying framework for structure and motion recovery from image sequences. *Proceedings of IEEE International Conference Computer Vision*, Boston, MA, pp. 314–320.

McLauchlan, P. 2000. A batch/recursive algorithm for 3D scene reconstruction. *Proceedings of IEEE Conference Computer Vision Pattern Recognition*, Vol. 2, Hilton Head, SC, pp. 738–743.

Milne, P. 1983. *Underwater Acoustic Positioning Systems*. Gulf Publishing Company, Houston, TX.

Neira, J. and Tardos, J. 2001. Data association in stochastic mapping using the joint compatibility test. *IEEE Transactions Robotics Automation* 17(6):890–897.

Newman, P. 1999. On the structure and solution of the simultaneous localisation and map building problem. PhD dissertation, Australian Centre for Field Robotics, The University of Sydney, March.

Nistér, D., Naroditsky, O., and Bergen, J. 2004. Visual odometry. *Proceedings of IEEE Conference Computer Vision Pattern Recognition*, Vol. 1, pp. 652–659.

Paskin, M. 2002 Thin junction tree filters for simultaneous localization and mapping. University of California, Berkeley, Technical Report CSD-02-1198, September.

Paskin, M. 2003. Thin junction tree filters for simultaneous localization and mapping. *Proceedings of International Joint Conference Artificial Intelligence*, San Francisco, CA, pp. 1157–1164.

Pizarro, O. 2004. Large scale structure from motion for autonomous underwater vehicle surveys. PhD dissertation, Massachusetts Institute of Technology /Woods Hole Oceaonographic Institution Joint Program, September.

Pizarro, O., Eustice, R., and Singh, H. 2003. Relative pose estimation for instrumented, calibrated imaging platforms. *Proceedings of Digital Image Computing Applications*, Sydney, Australia, December, pp. 601–612.

Pizarro, O., Eustice, R., and Singh, H. 2004. Large area 3D reconstructions from underwater surveys. *Proceedings of IEEE/MTS OCEANS Conference and Exhibition*, Vol. 2, Kobe, Japan, November, pp. 678–687.

Pollefeys, M., Gool, L. V., Vergauwen, M., Verbiest, F., Cornelis, K., Journal Tops, and Koch, R. 2004. Visual modeling with a hand-held camera. *International Journal Computer Vision* 59(3):207–232.

Repko, J. and Pollefeys, M. 2005. 3D models from extended uncalibrated video sequences: addressing key-frame selection and projective drift. *Proceedings of International Conference 3-D Digital Imaging and Modeling*, pp. 150–157.

Reynolds, J., Highsmith, R., Konar, B., Wheat, C., and Doudna, D. 2001. Fisheries and fisheries habitat investigations using undersea technology. *Proceedings of IEEE/MTS OCEANS Conference and Exhibition*, Vol. 2, Honolulu, HI, USA, November, pp. 812–820.

Roumeliotis, S., Johnson, A., and Montgomery, J. 2002. Augmenting inertial navigation with image-based motion estimation. *Proceedings of IEEE International Conference Robotics Automation*, Vol. 4, Washington, DC, May, pp. 4326–4333.

Rousseeuw, P. and Leroy, A. 1987. *Robust regression and outlier detection*. John Wiley and Sons, New York.

Se, S., Lowe, D., and Little, J. 2005. Vision-based global localization and mapping for mobile robots. *IEEE Transactions Robotics* 21(3):364–375.

Shewchuk, J. 1994. An introduction to the conjugate gradient method without the agonizing pain. Carnegie Mellon University, Technical Report CMU-CS-94-125, August.

Singh, H., Armstrong, R., Gilbes, F., Eustice, R., Roman, C., Pizarro, O. and Torres, J. 2004. Imaging coral I: imaging coral habitats with the SeaBED AUV. *Journal Subsurface Sensing Technology and Applications* 5(1):25–42.

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

Thrun, S., Koller, D., Ghahramani, Z., Durrant-Whyte, H., and Ng, A. 2002. Simultaneous mapping and localization with sparse extended information filters: theory and initial results. *Proceedings of International Workshop Algorithmic Foundations of Robotics*, J. Boissonnat, J. Burdick, K. Goldberg, and S. Hutchinson, Eds, Nice, France.

Thrun, S., Liu, Y., Koller, D., Ng, A., Ghahramani, Z., and Durrant-Whyte, H. 2004. Simultaneous localization and mapping with sparse extended information filters. *International Journal Robotics Research* 23(7-8):693–716.

Uhlmann, J. 1997. A culminating advance in the theory and practice of data fusion, filtering, and decentralized estimation. Covariance Intersection Working Group (CIWG), Technical Report, 1997.

Van Gool, L., Defoort, F., Hug, J., Kalberer, G., Koch, R., Martens, D., Pollefeys, M., Proesmans, M., Vergauwen, M., and Zalesny, A. 2000. Image-based 3D modeling: modeling from reality. *Proceedings of NATO Advanced Workshop on Confluence of Computer Vision and Computer Graphics*, ser. NATO Science Series, S. Ljubljana, A. Leonardis, F. Solina, and R. Bajcsy, Eds, Vol. 84, Kluwer Academic Publishers, pp. 161–178.

Whitcomb, L., Yoerger, D., and Singh, H. 1999. Combined Doppler/LBL based navigation of underwater vehicles. *Proceedings of International Symposium Unmanned Untethered Submersibles Technology*, Durham, New Hampshire, May.

Zhang, Z. 1998. Determining the epipolar geometry and its uncertainty: a review. *International Journal Computer Vision* 27(2): 161–198.