Pose-graph Visual SLAM with Geometric Model Selection for Autonomous Underwater Ship Hull Inspection

Ayoung Kim* and Ryan Eustice†
*Department of Mechanical Engineering
†Department of Naval Architecture & Marine Engineering
University of Michigan
Ann Arbor, Michigan 48109-2145
email:{ayoungk, eustice}@umich.edu

Abstract— **This paper reports the application of vision based simultaneous localization and mapping (SLAM) to the problem of autonomous ship hull inspection by an underwater vehicle. The goal of this work is to automatically map and navigate the underwater surface area of a ship hull for foreign object detection and maintenance inspection tasks. For this purpose we employ a pose-graph SLAM algorithm using an extended information filter for inference. For perception, we use a calibrated monocular camera system mounted on a tilt actuator so that the camera approximately maintains a nadir view to the hull. A combination of SIFT and Harris features detectors are used within a pairwise image registration framework to provide camera-derived relative-pose constraints (modulo scale). Because the ship hull surface can vary from being locally planar to highly three-dimensional (e.g., screws, rudder), we employ a geometric model selection framework to appropriately choose either an essential matrix or homography registration model during image registration. This allows the image registration engine to exploit geometry information at the early stages of estimation, which results in better navigation and structure reconstruction via more accurate and robust camera-constraints. Preliminary results are reported for mapping a 1,300 image data set covering a $30\,\mathrm{m}$ by $5\,\mathrm{m}$ section of the hull of a USS aircraft carrier. The post-processed result validates the algorithm's potential to provide in-situ navigation in the underwater environment for trajectory control, while generating a texture-mapped 3D model of the ship hull as a byproduct for inspection.**

I. Introduction

Present day means for ship hull and port security inspection require either putting divers in the water or piloting a remotely operated vehicle (ROV) over the area of interest—both of these options are manpower intensive and generally cannot quantitatively guarantee 100% survey coverage. There is a growing need to automate this task to allow autonomous robotic inspection of ships and port facilities for foreign objects such as limpet mines or improvised explosive devices (IEDs). Automating this task, however, is challenging. Current tethered robotic inspection systems present issues of snagging, maneuver degradation, and tether management, all of which make maneuvering around the ship difficult. Moreover, current robotic inspection methods require human in-the-loop intervention for both sensory interpretation and control (ROV pilot). Navigation feedback in these scenarios is typically performed using acoustic beacon time-of-flight ranging [1]. This requires setup and calibration of the beacon infrastructure, and therefore vitiates our ability to rapidly and repeatably inspect multiple underwater port structures.

In light of this, there exists a need to automate this task through the use of untethered autonomous underwater vehicles (AUVs). To do so requires overcoming several present-day science and technology challenges inherent to the inspection task. For example, areas around ships in pier are severely confined, cluttered, and complex sensing environments (e.g., acoustically, magnetically, optically). This necessitates the need for advanced navigation and localization systems that can work in confined, environmentally noisy spaces. Additionally, we would like the inspection method to quantitatively ensure 100% survey coverage of the hull, pier structures, and pilings for IED and foreign object detection. The underlying algorithm should facilitate in-situ sensor-reactive navigation while accommodating long-term map-based learning during revisited exploration (a prerequisite for hull change detection). Moreover, the increased diversity of threat objects and associated potential for false alarms in a cluttered environment require that fusion take place from multiple types of sensors for robustness and redundancy. In combination, all of these challenges and requirements, suggest that a SLAM based approach could hold the promise of accommodating the needs of autonomous automated hull search and inspection by AUVs.

A. Review of Underwater Vehicle Navigation Systems

The global positioning system (GPS) [2] has revolutionized navigation for land-based and airborne robotic missions; however, because seawater is opaque to most forms of electro-magnetic radiation, including GPS signals [3], determining position underwater must rely upon other means. A variety of operational alternatives exist to determine the position of an underwater vehicle, but these conventional solutions have limits.

Two broad categories of underwater navigation methods exist for localizing vehicles and instruments: absolute positioning and relative dead reckoning. The traditional long-baseline (LBL) method of underwater positioning estimates absolute position by measuring time-of-flight ranges to fixed beacons [4], [5]. The precision of this estimate is bounded,

and the accuracy is determined by system biases. The range of this solution is limited to a few kilometers in the best acoustic conditions and the positioning resolution is on the order of 1 meter. The slow update rate of LBL is constrained by the acoustic travel times—typically updating every few seconds. In contrast to slow, coarse, but absolute LBL positioning, a Doppler velocity log (DVL) or inertial navigation system (INS) instead estimates distance traveled to infer position. Dead-reckoning is fast (~ 10 Hz) and delivers fine resolution (~ 1 cm), but the precision of this relative measurement is unbounded, growing monotonically with time. This makes it difficult to return to a known location or to relate measurements globally to one another.

B. Underwater SLAM

Over the past decade, a significant research effort within the terrestrial mobile robotics community has been to develop SLAM-based navigation algorithms that eliminate the need for additional infrastructure and bound position error growth to the size of the environment—a key prerequisite for truly autonomous navigation. The goal of this work has been to exploit the perceptual sensing capabilities of robots to correct for accumulated odometric error by localizing the robot with respect to landmarks in the environment [6], [7].

Two of the major challenges of the SLAM problem is *(i)* defining fixed features from raw sensor data and *(ii)* establishing measurement to feature correspondence (i.e., the problem of data association [8]). Both of these tasks can be nontrivial—especially in an unstructured underwater environment. In man-made environments, typically composed of planes, lines and corners, point features can be more easily defined; however, complex underwater environments pose a more challenging task for feature extraction and matching.

One SLAM methodology that has seen recent success in the near seafloor underwater realm is to apply a pose-graph scan-matching approach, as reported in [9]–[12]. Pose-graph SLAM approaches do not require an explicit representation of features and instead use a data-driven approach based upon extracting relative-pose constraints. These techniques have seen good success when applied to an unstructured seafloor environment. The main idea behind this methodology is that registering overlapping perceptual data, for example optical imagery as reported in [11] or sonar bathymetry as reported in [12], introduces spatial drift-free edge constraints into the pose-graph. These spatial constraints effectively allow the robot to close the loop when revisiting a previously visited place, thereby resetting any accumulated dead-reckoning error.

II. Visually Augmented Navigation

In this section we briefly review our pose-graph monocular SLAM formulation, which we call visually augmented navigation (VAN). We refer the reader to [13] and [14] for more details.

A. State Representation

We model the vehicle state using a 6-degree of freedom (DOF) representation for pose (position and Euler orientation), $\mathbf{x}_v = [x, y, z, \phi, \theta, \psi]^\top$, where pose is defined in a local-navigation frame with respect to the hull of the ship. A continuous-time constant-velocity kinematic model driven by white noise is used to approximate the time-evolution of vehicle state. This is then linearized and discretized to provide a linear time-varying discrete-time model of the vehicle dynamics.

In our work, we employ a pose-graph SLAM representation of the environment and therefore augment our state description to include a collection of historical vehicle poses sampled at regular spatial intervals throughout the environment. Each of these pose samples, $\mathbf{x}_{v_i}$, corresponds to the time instance t_i of a key frame stored by our visual perception process. Therefore, our augmented state representation is expressed as follows for n key frames

$$\boldsymbol{\xi}^\top = \left[\mathbf{x}_{v_1}, \cdots, \mathbf{x}_{v_i}, \cdots, \mathbf{x}_{v_n}, \mathbf{x}_v(t)\right].$$

The distribution of this augmented state representation is jointly Gaussian and parameterized in the inverse covariance form as

$$\boldsymbol{\eta} = \Lambda\boldsymbol{\mu} \;\text{ and }\; \Lambda = \Sigma^{-1}$$

where $\boldsymbol{\eta}$ and Λ are the information vector and matrix, respectively, as written in terms of the more familiar $\boldsymbol{\mu}$, Σ mean and covariance parameterization.

B. Camera Constraints

Pairwise registration of key frames acquired by the calibrated monocular vision system results in relative-pose constraints modulo scale between historical pose samples in $\boldsymbol{\xi}$. Here, the 5-DOF camera measurement $\mathbf{z}_{ji}$ between poses $\mathbf{x}_{v_i}$ and $\mathbf{x}_{v_j}$ is modeled as an observation of the azimuth α_{ji} and elevation angle β_{ji} of the baseline direction of motion, and the relative Euler orientation $\phi_{ji}, \theta_{ji}, \psi_{ji}$ between the two poses.

$$\mathbf{z}_{ji} = \mathbf{h}_{ji}(\mathbf{x}_{v_j}, \mathbf{x}_{v_i}) = \left[\alpha_{ji}, \beta_{ji}, \phi_{ji}, \theta_{ji}, \psi_{ji}\right]^\top$$

The Jacobian of $\mathbf{z}_{ji}$ with respect to $\boldsymbol{\xi}$ is sparse

$$\mathrm{H}_{\mathbf{x}} = \begin{bmatrix} 0 & \cdots & \frac{\partial \mathbf{h}_{ji}}{\partial \mathbf{x}_{p_j}} & \cdots & 0 & \frac{\partial \mathbf{h}_{ji}}{\partial \mathbf{x}_{p_i}} & \cdots & 0 \end{bmatrix},$$

which preserves sparsity in the information matrix, Λ, during measurement updates.

C. Extended Information Filter Inference

For the Gaussian distribution, the extended information filter (EIF) provides efficient inference when using a pose-graph representation due to the exact mathematical sparsity imposed by the pose-graph state representation [16]. The reader is referred to publications [14], [16] for details regarding the EIF motion prediction, measurement update, and state augmentation equations as they pertain to EIF inference with the information vector and matrix parameterization.

The EIF is the dual of the extended Kalman filter (EKF), and therefore yields an equivalent result to the pose-graph

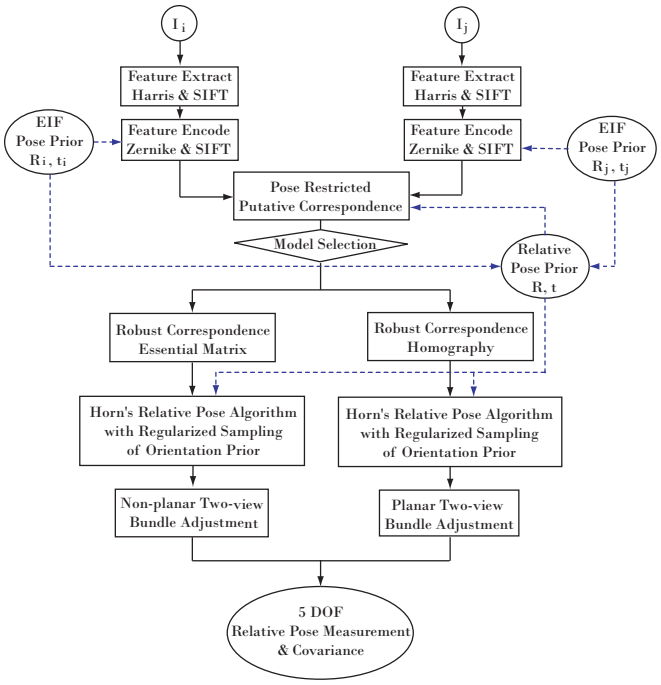


Fig. 1. Overview of VAN's pairwise image registration engine. Prior knowledge of relative-pose $(\mathrm{R}, \mathbf{t})$ comes from the EIF state estimate and is used in the two-view registration framework where applicable. For each image, both SIFT and Harris feature detectors are used for extraction. Putative correspondences are established based upon a pose-constrained correspondence search [13]. Following the inlier set detection, Horn's algorithm [15] is used to obtain an initial estimate of relative-pose, which is then used as an initial guess in the final two-view bundle adjustment step to obtain the optimal 5-DOF camera constraint.

EKF; however, the EIF does not exhibit the quadratic computational complexity of the EKF, allowing it to approach near-linear inference time.

III. Geometric Model Selection

The process of motion estimation from a pair of images in the VAN framework involves three steps: 1) detecting and writing descriptors for features in each image; 2) establishing inlier correspondences and fitting a geometric model using a robust estimation framework; and 3) optimizing the camera relative-pose constraint. In the first step, feature points are extracted from the pair of images and encoded into description vectors. These are subsequently used to establish putative correspondence matches via an appearance similarity metric. Then, in the second step, an inlier correspondence set is determined using a robust RANSAC model fitting framework. In this step, geometric model selection is used to provide accurate correspodences. Finally, the third step involves two-view bundle adjustment to find the optimal value of the camera-derived relative-pose constraint. The overall algorithm is depicted in Fig. 1.

A. Feature Descriptor

A combination of SIFT [17] and Harris [18] image features are used for feature extraction. For encoding, we use SIFT keys with 128 vector key descriptors for the SIFT interest points, and Zernike moments [19] for a local patch around the Harris interest points. By changing the similarity threshold of the Zernike feature vector, we adjust the putative correspondence ratio of SIFT and Harris points, using SIFT as the major feature extraction algorithm assisted by Harris.

B. Model Selection

The ship hull surface can vary from being locally piecewise planar to highly three dimensional (e.g., rudders and propeller). Because of this variance, we adopt a model selection framework to appropriately choose either a homography or essential matrix image registration model during the pairwise registration step.

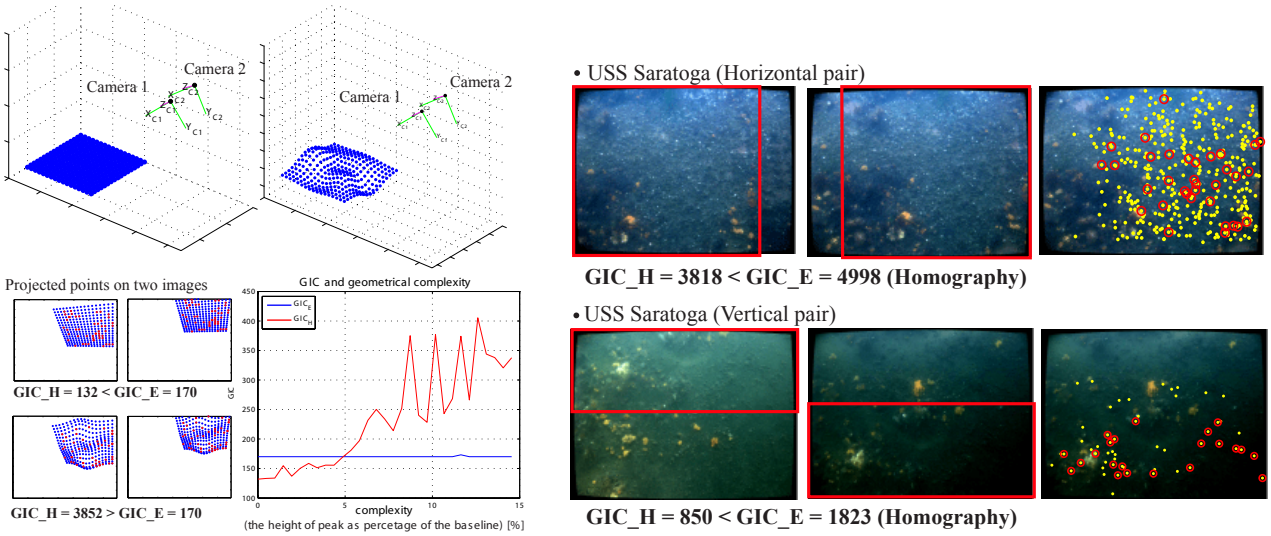
There are several reported algorithms in the literature aimed at overcoming registration ambiguity by using a model-selection framework to automatically determine the appropriate image registration model during registration. In the work of [20]–[23], several different metrics such as Akaike Information Criterion (AIC) [20], Geometric AIC [21], Robust AIC [22] and Geometric Information Criterion (GIC) [23] were proposed in order to determine the geometric complexity of the image scene, and were evaulated in [24]. These metrics are used to determine the correct registration model by measuring the summation of the reprojection error and a model complexity penalty. The algorithm choose the model with smaller metric as the correct registration model. In this paper, we have adopted Torr's GIC [23],

$$GIC = \sum \frac{e_i^2}{\sigma^2} + \lambda_1 dN + \lambda_2 P$$

Here, e_i is the i^{th} correspondence pixel reprojection error, N is the number of correspondences, d is the dimension of the registration manifold (2 for homography and 3 for essential matrix), P is the number of parameters of the registration model (8 for homography and 5 for essential matrix) and $\lambda_1 = 2$ and $\lambda_2 = 4$ are penalty weights as recommended in the literature.

In two key respects, the homography model is preferable to the essential matrix model. One key benefit of selecting the homography model is that using it reduces the number of variables involved and thereby simplifies the optimization problem. Given N number of the correspondences, the total number of the variables in the two-view bundle adjustment is $2N + 8$ for the homography model, and $3N + 5$ for the essential matrix model. A further advantage to the homography is that, in planar structures, it provides greater accuracy in correspondences than does the essential matrix model. As can be seen in Fig. 2(b), the essential matrix model, in all areas apart from the overlap region, may lead to false correspondences because that model measures the geometric distance of a point from the epipolar line whereas the homography model measures the point to point geometric error.

The homography model is, however, not feasible in a complex structure due to the large reprojection error. Fig. 2(a) presents the results from testing this criteria with a synthetic data set of varying complexity. Since the GIC for homography is smaller for the synthetic planar data,



(a) Synthetic Data

(b) Real Images

Fig. 2. Geometric model selection metric for a synthetic (a) and real data set(b). In the synthetically projected images in (a), the blue dots represent the projected structure points in each image, while the red dots indicate the randomly sampled true correspondences as seen from both cameras. The synthetic data has been generated using MATLAB peaks function by increasing the height of the peak. In the real imagery (b), the left and middle columns show two overlapping image pairs while the right column shows the inlier correspondences. Indicated in yellow are correspondences that satisfy the epipolar geometry constraint, and shown in red are those that are consistent with the homography constraint. The red rectangles indicate the overlap area between the two images.

it was chosen as an appropriate model for the data. As the structure becomes complex and non-planar, the GIC for the homography becomes increasingly larger than that of the essential matrix, while the reprojection error from the essential matrix does not greatly vary commensurate with the complexity of the structure. As shown in the graph in Fig. 2(a), this non-varying level of error keeps the GIC value constant despite increasing complexity, while this error increases continuously with the homography model.

By solving this homography plane detection problem in the VAN algorithm, this model selection framework allows the image registration algorithm to exploit geometric information at an early stage of estimation before the two-view optimization step. This results in better navigation and structure reconstruction via more accurate and robust camera-derived pose-graph constraints.

C. Two-view Bundle Adjustment

The final step of the image registration engine is to optimally recover the relative-pose constraint derived from the camera correspondence pairs. Here, the objective function varies according to the model selection metric we determined in the previous section.

1) General structure: When the model selection criteria results in selecting the essential matrix epipolar geometry constraint, the optimal camera-derived relative-pose constraint is found using a two-view bundle adjustment. Bundle adjustment minimizes the reprojection error in both images by optimizing camera relative-pose (i.e., $\mathrm{R}, \mathbf{t}$) and triangulated 3D structure (i.e., $\mathbf{X}_n$). The camera projective matrices are $\mathrm{P}_i = \mathrm{K}[\mathrm{I} \,|\, \mathbf{0}]$ and $\mathrm{P}_j = \mathrm{K}[\mathrm{R} \,|\, \mathbf{t}]$ where the camera internal parameters, K, are the same and known from calibration.

The two-view bundle adjustment objective function $f(.)$ is the sum squared reprojection error in the two images taken over all N common correspondences, i.e.,

$$\min_{\mathrm{R},\mathbf{t},\mathbf{X}_n} f(\mathrm{R}, \mathbf{t}, \mathbf{X}_n) = \sum_{n=1}^{N} \mathbf{e}_n^\top \mathbf{e}_n$$

where

$$\mathbf{e}_n = \begin{bmatrix} \underline{\mathbf{u}}_{i_n} - \mathrm{P}_i \underline{\mathbf{X}}_n \\ \underline{\mathbf{u}}_{j_n} - \mathrm{P}_j \underline{\mathbf{X}}_n \end{bmatrix}.$$

2) Planar structure: When the geometric model selection criteria chooses the homography registration model, the objective function is chosen to minimize the sum squared reprojection error using a plane induced homography registration model. The optimization is performed over the plane induced homography parameters (R, $\mathbf{t}$, $\mathbf{n}$, d) and the optimal image correspondences $\hat{\mathbf{u}}_{i_n}$ that satisfy the homography mapping exactly (i.e., $\underline{\hat{\mathbf{u}}}_{j_n} = \mathrm{H}\underline{\hat{\mathbf{u}}}_{i_n}$).

The plane induced homography H is written in terms of the world plane $\pi = [\mathbf{n}^\top, d]^\top$ and camera relative-pose R, $\mathbf{t}$ as [25]

$$\mathrm{H} = \mathrm{K}(\mathrm{R} - \mathbf{t}\mathbf{n}^\top / d)\mathrm{K}^{-1}.$$

Here, the world plane normal $\mathbf{n}$ and orthogonal distance d expressed in the camera i reference frame.

Given the correspondence pairs between two images, the optimization solves for the optimal value of homography parameters together with the image points, i.e.,

$$\min_{\mathrm{R},\mathbf{t},\mathbf{n},d,\underline{\hat{\mathbf{u}}}_{i_n}} f(\mathrm{R}, \mathbf{t}, \mathbf{n}, d, \underline{\hat{\mathbf{u}}}_{i_n}) = \sum_{n=1}^{N} \mathbf{e}_n^\top \mathbf{e}_n$$

where

$$\mathbf{e}_n = \begin{bmatrix} \underline{\mathbf{u}}_{i_n} - \underline{\hat{\mathbf{u}}}_{i_n} \\ \underline{\mathbf{u}}_{j_n} - \mathrm{H}\underline{\hat{\mathbf{u}}}_{i_n} \end{bmatrix}.$$

IV. Experimental Results

In May 2008, over 1,300 underwater images of the aircraft carrier USS Saratoga (Fig. 3) were collected at AUVFest2008 in Newport, Rhode Island.[1] The experiment was done in collaboration with MIT and Bluefin Robotics using their autonomous underwater hull inspection vehicle, the HAUV (Hovering Autonomous Underwater Vehicle) [26]. The HAUV is equipped with a Teledyne 1200 kHz RDI Explorer DVL, depth sensor, fiber-optic gyro, and Didson sonar. For the experiment, we designed a strap on camera and light system consisting of a Prosilica GC1380C 12-bit high resolution monocular camera and incandescent underwater light. The experiment consisted of seven ~ 30 m legs of a boustrophedon survey, each spaced 0.5 m apart in depth. The camera and DVL were mounted on tilt actuators so that they approximately maintained a nadir view to the hull. The standoff position of the robot was controlled at 1.5 m from the hull throughout the experiment with a horizontal trajectory speed of 0.5 m/s. Six bucket shaped targets were attached to hull to validate the ability to inspect the ship hull and navigate around it. (These targets were not used by the visual SLAM algorithm and were installed as acoustic features for sonar detection.)

The final trajectory resulting from visual SLAM with model selection is shown in Fig. 5(a). The red and green links in the figure indicate the camera constraints in the pose-graph; the green links indicate sequential pose-constraints coming from the registration of temporally consecutive image pairs, and the red lines indicate pose constraints derived from non-temporally sequential image pairs. The camera was restarted while the robot was returning to the bow in the second leg of the trajectory. This resulted in a blank section in the middle of the second trackline during which the uncertainty ellipsoid inflated as the robot moved without camera measurements over this section. However, once the camera was restarted at the end of the second leg, the uncertainty was reduced as the VAN algorithm re-localized the vehicle by adding additional pose constraints to the previous track (first leg). This increase and decrease of the uncertainty is repeated as the robot hovers and moves from the featureless region (toward stern right) to the feature-rich region (toward bow left).

Fig. 4 shows the change in pose uncertainty versus the total path length that the underwater robot traveled. The uncertainty of the robot pose using dead-reckoned (DR) navigation is the blue line, and shows monotonically increasing uncertainty along the total path as it localizes itself based upon DVL-based integration. As can be seen from the red line, camera-constraints used in the VAN algorithm significantly constrain the uncertainty, preventing it from monotonically increasing. The sawtooth pattern in the VAN uncertainty estimate is the result of navigating into the visually featureless region near the stern, and then returning back toward the more feature-rich bow area where image registration was more repeatable.

[1] AUVFest is an Office of Naval Research sponsored AUV field demonstration event.

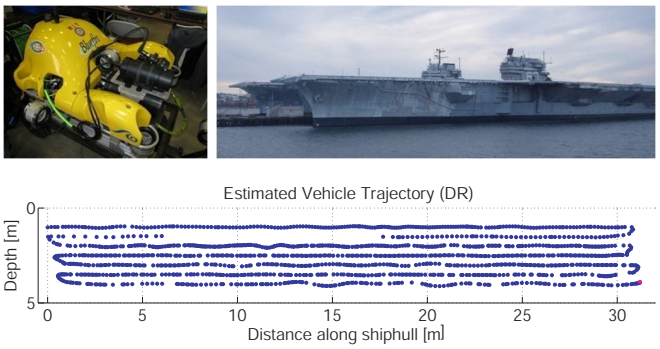


Fig. 3. Hull inspection experimental setup. (top left) Bluefin Robotics HAUV-1B. (top right) USS Saratoga aircraft carrier. (bottom) Hull-relative trajectory estimated from odometry (DVL).

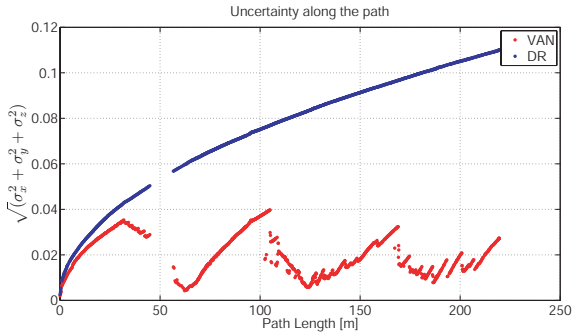


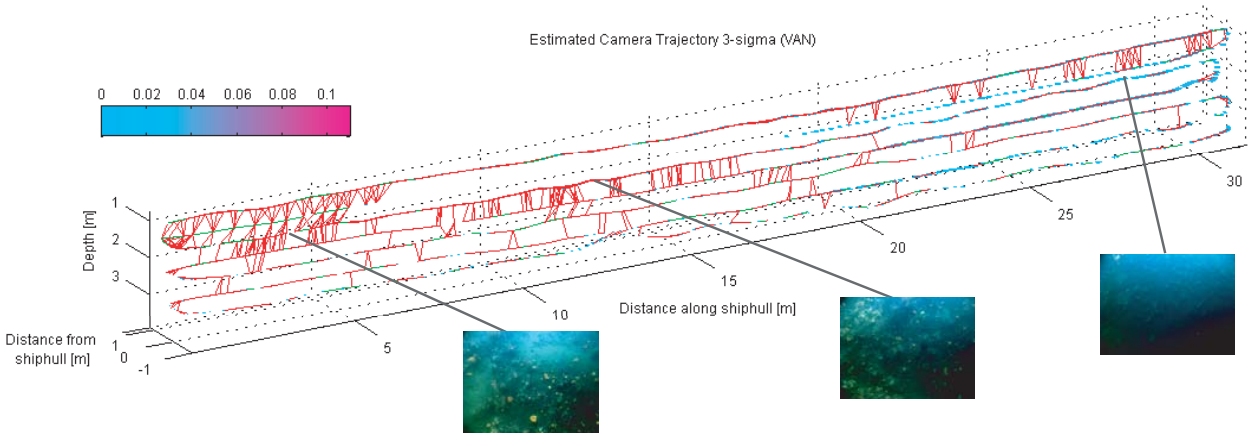
Fig. 4. Uncertainty plot for overall path. Whereas the uncertainty of DR navigation (blue line) is increasing monotonically, the uncertainty of VAN is bounded (red line).

As a byproduct of VAN trajectory estimation, the 3D structure of the ship hull can be reconstructed using the final trajectory estimate and the pairwise image correspondences. Surface fitting with texture mapping was conducted to generate a 3D photomosaic (Fig. 6(b)). All the targets were found as indicated by the six white circles, which appear in the texture-mapped reconstruction.

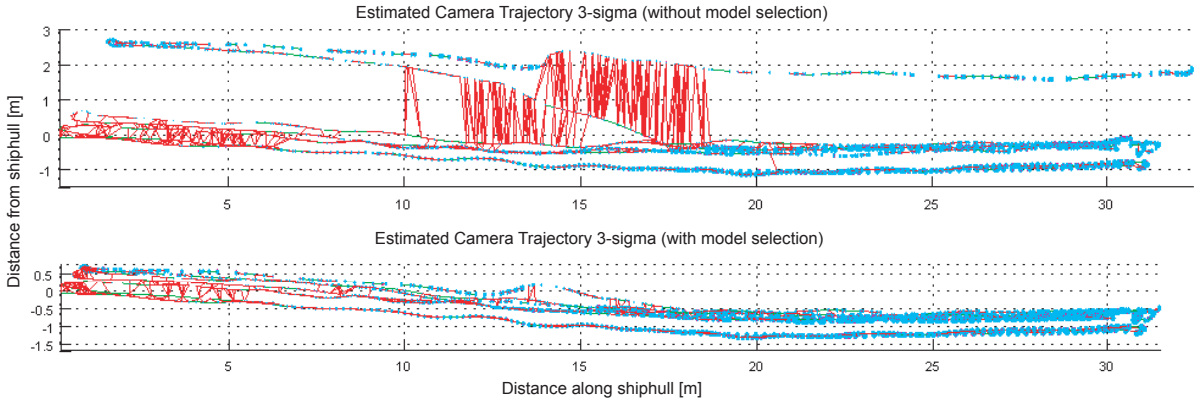
A movie showing a virtual fly-through over the 3D photomosaic is available as multimedia data `saratoga_3d_photomosaic.mp4`.

V. Conclusion

This paper reported an application of visual underwater SLAM to the task of autonomous underwater ship hull inspection. We presented an extension to the VAN algorithm that uses geometrical model selection providing accurate correspondences in the image-based motion estimation. This increases the robustness and accuracy of the camera-derived pose constraint since a large portion of the hull is locally flat. We presented results for mapping a 30 m by 5 m section of the hull. While we do not have independent ground-truth to validate our trajectory estimate, we note that the recovered



(a) 3D trajectory estimate using VAN algorithm



(b) Top view of the trajectory

Fig. 5. The VAN estimated trajectory and reconstructed structure with texture mapping. The camera-derived pose constraints are shown in (a) with red and green links. Each vertex in (a) represents a node in the pose-graph enclosed by its 3σ covariance uncertainty ellipsoid. Due to the change of the visual feature richness on the hull, the uncertainty ellipsoid inflates when the vehicle is not able to build enough pose constraints, but then deflate once VAN creates camera constraints with previous tracklines. Three small figure insets in (a) depict the typical feature richness for different regions of the hull. Figure (b) shows the top-down view of the vehicle trajectory along the ship hull. The upper plot in (b) shows the trajectory result without model selection, while the lower plot shows the trajectory with model selection. Note that without model selection, the estimated trajectory is mislead with 2 meters of offset due to the bad essential matrix correspondences making it through RANSAC.

trajectory results in a smooth surface reconstruction indicating that the recovered poses must be highly self-consistent.

Our next research focus is in the area of perception driven control, which will couple the SLAM navigation estimate back into the trajectory generation for an optimal mapping policy. As shown in the current data set, large areas of the hull survey occur in nearly feature-less regions, while other regions exhibit a high density of features. For these scenarios, it would be ideal to have the vehicle autonomously return to a previously known feature rich area in an intelligent way whenever its pose uncertainty grows too high. In this manner, the vehicle will be able to maintain an upper bound on its pose uncertainty at any point along the hull, and thereby quantitatively guarantee 100% survey coverage for inspection tasks.

ACKNOWLEDGMENTS

This work is supported through a grant from the Office of Naval Research (ONR) (Award #N00014-07-1-0791).

REFERENCES

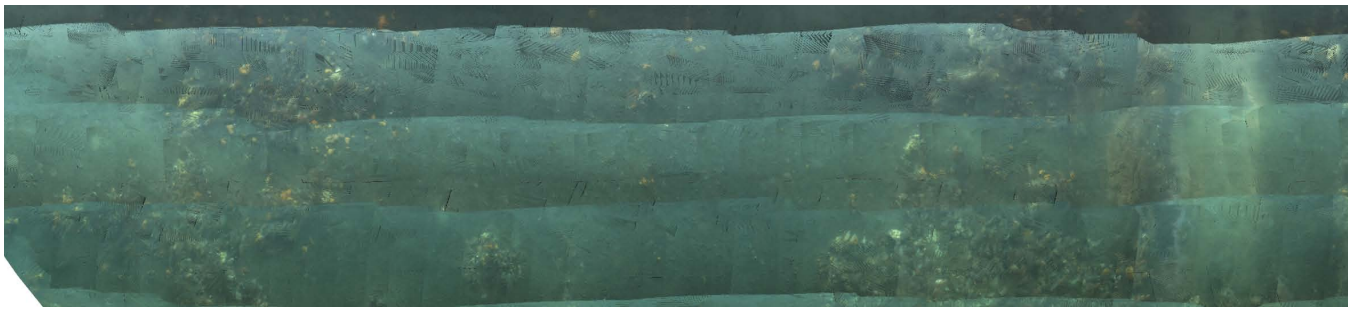
[1] S. Smith and D. Kronen, "Experimental results of an inexpensive short baseline acoustic positioning system for AUV navigation," in Proceedings of the IEEE/MTS OCEANS Conference and Exhibition, vol. 1, Halifax, NS, 1997, pp. 714–720.

[2] B. Hofman Wellenhof, H. Lichtenegger, and J. Collins, Global positioning system (GPS): theory and practice, 5th ed. New York: Springer-Verlag, 2001.

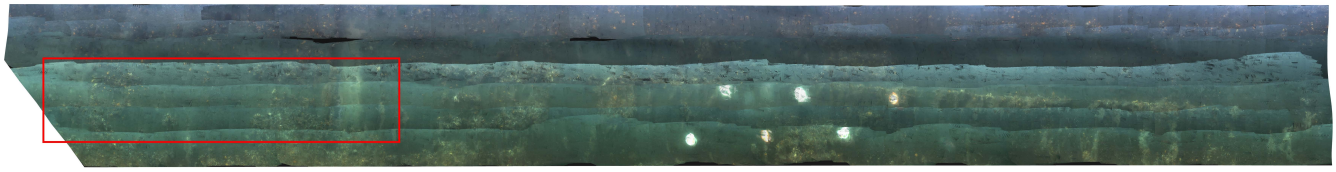
[3] W. Stewart, "Remote-sensing issues for intelligent underwater systems," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Maui, HI, USA, 1991, pp. 230–235.

[4] P. Milne, Underwater acoustic positioning systems. Houston: Gulf Publishing Company, 1983.

[5] M. Hunt, W. Marquet, D. Moller, K. Peal, W. Smith, and R. Spindel, "An acoustic navigation system," Woods Hole Oceanographic Institution, Tech. Rep. WHOI-74-6, Dec. 1974.



(a) Zoomed view of inset



(b) Texture mapped reconstruction

Fig. 6. Triangulated 3D points can be fitted to obtain a smooth surface reconstruction, and texture mapped to create a 3D photomosaic as in (b). The six white dots that appear in (b) are the targets used for visual verification of the hull inspection utility of the algorithm.

[6] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: part I," IEEE Robotics and Automation Magazine, vol. 13, no. 2, pp. 99–110, Jun. 2006.

[7] T. Bailey and H. Durrant-Whyte, "Simultaneous localization and mapping (SLAM): part II." IEEE Robotics and Automation Magazine, vol. 13, no. 3, pp. 108–117, Sep. 2006.

[8] J. Neira and J. Tardos, "Data association in stochastic mapping using the joint compatibility test," IEEE Transactions on Robotics and Automation, vol. 17, no. 6, pp. 890–897, Dec. 2001.

[9] S. Fleischer, "Bounded-error vision-based navigation of autonomous underwater vehicles," Ph.D. dissertation, Stanford University, May 2000.

[10] R. Garcia, J. Batlle, X. Cufi, and J. Amat, "Positioning an underwater vehicle through image mosaicking," in Proceedings of the IEEE International Conference on Robotics and Automation, vol. 3, Seoul, May 2001, pp. 2779–2784.

[11] R. M. Eustice, R. Camilli, and H. Singh, "Towards bathymetry-optimized Doppler re-navigation for AUVs," in Proceedings of the IEEE/MTS OCEANS Conference and Exhibition, Washington D.C., Sep. 2005, pp. 1430–1436.

[12] C. Roman, "Self consistent bathymetric mapping from robotic vehicles in the deep ocean," Ph.D. dissertation, Massachusetts Institute of Technology / Woods Hole Oceaonographic Institution Joint Program, Jun. 2005.

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

[14] R. M. Eustice, H. Singh, J. J. Leonard, and M. R. Walter, "Visually mapping the RMS Titanic: conservative covariance estimates for SLAM information filters," International Journal of Robotics Research, vol. 25, no. 12, pp. 1223–1242, 2006.

[15] B. Horn, "Relative orientation," International Journal of Computer Vision, vol. 4, no. 1, pp. 59–78, Jan. 1990.

[16] R. M. Eustice, H. Singh, and J. J. Leonard, "Exactly sparse delayed-state filters for view-based SLAM," IEEE Transactions on Robotics, vol. 22, no. 6, pp. 1100–1114, Dec. 2006.

[17] M. Brown and D. Lowe, "Invariant features from interest point groups," in Proceedings of the British Machine Vision Conference, 2002, pp. 656–665. [Online]. Available: citeseer.ist.psu.edu/brown02invariant.html

[18] C. Harris and M. Stephens, "A combined corner and edge detector," in Proc. 4th Alvey Vision Conf., Manchester, U.K., 1988, pp. 147–151.

[19] P. Hew, "Geometric and Zernike moments," The University of Western Australia, Department of Mathematics, Tech. Rep., Oct. 1996.

[20] H. Akaike, "A new look at the statistical model identification," Automatic Control, IEEE Transactions on, vol. 19, no. 6, pp. 716–723, 1974. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1100705

[21] K. Kanatani and Y. Kanazawa, "Stabilizing image mosaicing by the geometric aic," in Proc. 2nd Workshop on Information-Based Induction Sciences, 1999, pp. 181–186.

[22] P. H. S. Torr, "Geometric motion segmentation and model selection," Phil. Trans. Royal Society of London A, vol. 356, pp. 1321–1340, 1998.

[23] P. Torr, "Model selection for two view geometry: A review," Microsoft Research, USA, Microsoft Research, pp. 277–301, 1998.

[24] N. Gheissari and A. Bab-Hadiashar, "Detecting cylinders in 3d range data using model selection criteria," 3D Digital Imaging and Modeling, International Conference on, vol. 0, pp. 158–163, 2005.

[25] R. Hartley and A. Zisserman, Multiple view geometry in computer vision. Cambridge University Press, 2000.

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