Advances in High Resolution Imaging from Underwater Vehicles

Hanumant Singh, Chris Roman, Oscar Pizarro, Ryan Eustice
Dept of Applied Ocean Physics and Engineering
Woods Hole Oceanographic Institution
Woods Hole, MA, USA

Abstract

Large area mapping at high resolution underwater continues to be constrained by the mismatch between available navigation as compared to sensor accuracy. In this paper we present advances that exploit consistency and redundancy within local sensor measurements to build high resolution optical and acoustic maps that are a consistent representation of the environment.

We present our work in the context of real world data acquired using Autonomous Underwater Vehicles (AUVs) and Remotely Operated Vehicles (ROVs) working in diverse applications including shallow water coral reef surveys with the Seabed AUV, a forensic survey of the RMS Titanic in the North Atlantic at a depth of 4100 meters using the Hercules ROV and a survey of the TAG hydrothermal vent area in the mid-Atlantic at a depth of 2600m using the Jason II ROV.

Specifically we focus on the related problems of Structure from Motion and Visually Based Navigation from underwater optical imagery assuming pose instrumented calibrated cameras. We present general wide baseline solutions for these problems based on the extension of techniques from the SLAM, photogrammetric and the computer vision communities. We also examine how such techniques can be extended for the very different sensing modality and scale associated with multi-beam bathymetric mapping. For both the optical and acoustic mapping cases we also show how the consistency in mapping can be used not only for better mapping but also to refine navigation estimates.

Introduction

A number of oceanographic applications require large area site surveys from underwater imaging platforms. Such surveys are typically required to study hydrothermal vents and spreading ridges in geology[1] , ancient shipwrecks and settlements in archaeology[2], forensic studies of modern shipwrecks and airplane accidents[3][4], and surveys of benthic ecosystems and species in biology[5]. Scientific users in these disciplines often rely on multiscalar, multisensor measurements to best characterize the environment.

At finer scales, for resolutions down to millimeters, optical imaging of the seafloor offers scientists a high level of detail and ease of interpretation. However light underwater suffers from significant attenuation and backscatter, limiting the practical coverage of a single image to a few square meters. To cover larger areas of interest, hundreds or thousands of images may be required. The rapid attenuation of the visible spectrum in water implies that a composite view of a large area (or photomosaic) can only be obtained by exploiting the redundancy in multiple overlapping images distributed over the scene. Although there has been considerable effort in this regard for land-based applications, the constraints on imaging underwater are far different. Mosaicing assumes that images come from an ideal camera (with compensated lens distortion) and that the scene is planar. Under these assumptions the camera motion will not induce parallax and therefore no 3D effects are involved and the transformation between views can then be correctly described by a 2D homography. These assumptions often do not hold in underwater applications since light attenuation and backscatter rule out the traditional land-based approach of acquiring distant, nearly orthographic imagery. Underwater mosaics of scenes exhibiting significant 3D structure usually contain significant distortions. In contrast to mosaicing, the information from multiple underwater views can be used to extract structure and motion estimates using ideas from structure from motion (SFM) and photogrammetry.

For coarser resolutions (O(10cm)), but covering far greater (O(10m-100m)) swaths, acoustic sensing centered at several hundred kilohertz is the modality of choice. Multibeam sensors mounted on underwater platforms can provide high resolution three dimensional scans of the environment that can be transformed into bathymetric maps.

Unfortunately for both optical and acoustic sensors, the fundamental limitation in converting high resolution sensor measurements into quantitative maps is the mismatch between sensor accuracy and navigation as illustrated schematically in Figure 1. Due to the rapid attenuation of the electromagnetic spectrum, GPS signals are not available underwater. Instead underwater imaging platforms typically rely on a combination of acoustic transponders and inertial navigation systems. Acoustic transponders[6], like sonar systems, must trade off range for resolution. Although transponders have been built to work as high as 300kHz providing centimeter level accuracy over an area of 100 square meters, the typical large area surveys utilize lower frequency (8-13 kHz) long-baseline transponders that provide meter level accuracy across several kilometers. The deployment of such systems is nontrivial and usually requires significant time and effort as each individual transponder must be deployed and its position independently calibrated.

Inertial navigation sensors such as doppler velocity logs used in combination with fiber optic or ring laser gyros can provide navigation estimates underwater[7] that grow as a function of time (distance traveled). However, such systems inherently provide an estimate whose error characteristic grows without bound over time (distance). Although expensive, from a cost, power and size standpoint, these systems are far easier to use as they are integral to the underwater vehicle and as such do not require any extra effort for deployment and use.

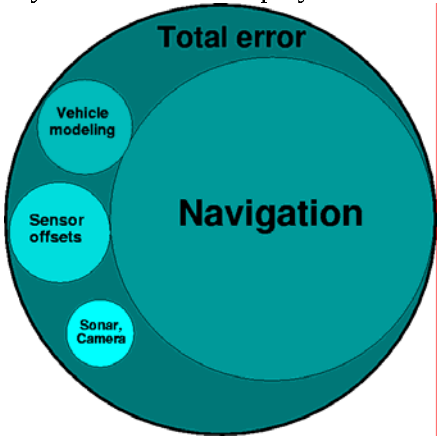

Figure 1. A schematic of error sources for high resolution optical and acoustic deep water mapping. Vehicle based mapping is navigationally limited in comparison to other potential error sources.

Structure from Motion Underwater – The two view Case

As outlined above, the fundamental problem of obtaining a large area perspective of an underwater scene is constrained by the attenuation and backscatter of light, the highly unstructured nature of underwater terrain, and issues associated with moving lighting on underwater robotic vehicles.

Our methodology for Structure from Motion takes a local to global approach inspired by mosaicing and other land-based applications of SFM [8][9][10][11][12] but differs in that it takes advantage of navigation and attitude information[13]. Local sequences are derived independently[11] and then registered in a global frame for bundle adjustment[10]. Our approach seems more suitable than pure sequential methods[9][11] because in an underwater survey each

3D feature appears only in a few images making the global solution look more like a series of weakly correlated local solutions.

We relate images using a feature-based approach under wide-baseline imaging conditions with changing illumination and unknown scene structure. A modified Harris corner detector yields interest points by selecting local maxima of the smaller eigenvalue of the second moment matrix. We extract features by determining a neighborhood around each interest point that is invariant to affine geometric transformations. In essence, we sample the neighborhood along lines radiating from the interest point. For each line we select the extrema of an affine invariant function (maximum difference in intensities between the interest point and points along the ray). The set of these maximal points defines the boundary of a region that can be extracted under affine geometric transformations. This region is approximated with an elliptical neighborhood which is then mapped onto the unit circle. These circular patches are normalized for affine photometric invariance. Features are then represented compactly using moment-based descriptors. We chose to use Zernike[13] moments as descriptors for their compactness and highly discriminative nature.

The core of the algorithm for SFM is based on robust estimation of the essential matrix[8]. Similarity of descriptor vectors is used to propose correspondences between features. The navigation-based estimates of inter-image motion and vehicle altitude are used to limit possible correspondences by propagating pose and altitude uncertainties through the two view point-transfer equation[13] as shown in Figure 2. A modified version of RANSAC determines the correspondences which are consistent with that essential matrix and the essential matrix consistent with the inliers as illustrated in Figure 3. In cases of multiple valid solutions we select the one closest (in the Mahalanobis distance sense) to the navigation-based prior. The inliers and the essential matrix estimate are used to produce a maximum a posteriori estimate of relative pose with the navigation-based estimates as a prior. The solution includes the triangulated 3D features.

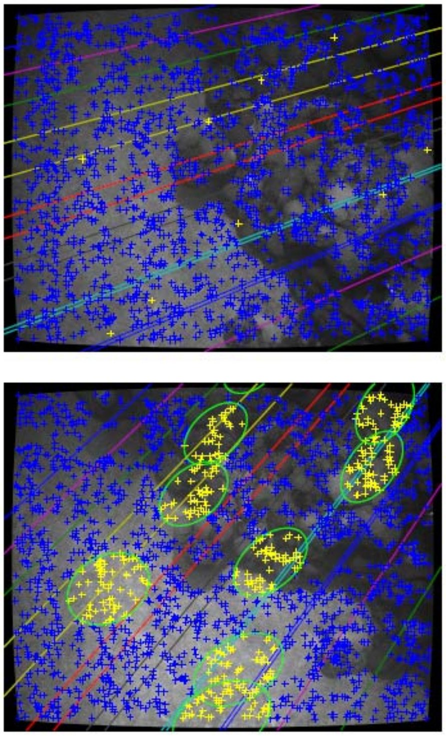

Figure 2. Prior pose restricted correspondence search on a pair of underwater coral reef images. A sampling of interest points are shown in the top image along with their color coded sensor instantiated epipolar lines. The bottom image shows the corresponding color coded constrained search regions for the interest points in the top image; the sensor instantiated epipolar lines; and the candidate interests points which fall within these regions.

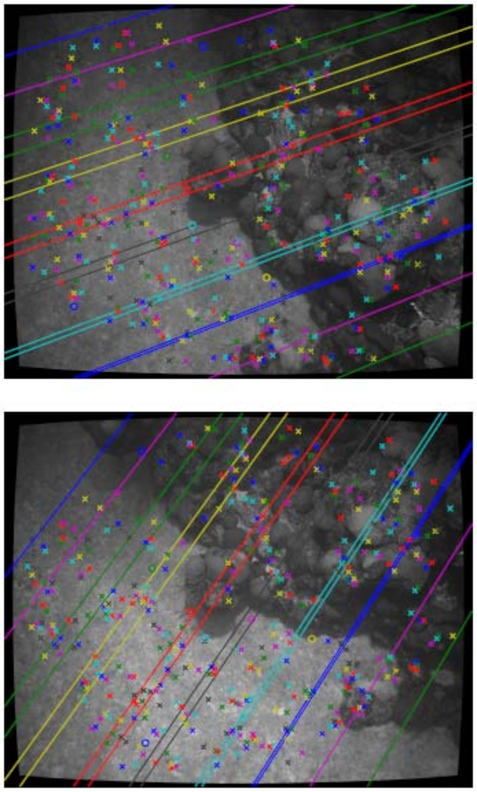

Figure 3. Epipolar geometry and correspondences. The given image pair illustrates the maximum likelihood refined image-based epipolar geometry. RANSAC determined 106 consistent inliers designated ‘x’, from the putative set of 116 matches. The rejected outliers are designated ‘o’.

Large Area Structure from Motion

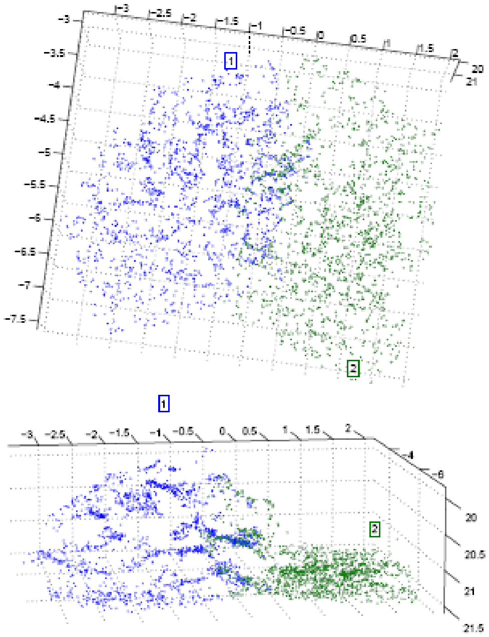

Figure 4. Two views of the registered submaps derived from images sequences that correspond to two neighbouring sections along the images shown in figures 2 and 3. The blue and green dots correspond to features from the neighbouring tracklines that have been successfully co-registered.

The temporal sequence of images is processed into a set of 3D submaps with estimates of coordinate transformations between temporally adjacent submaps (Figure 4). This can be viewed as a graph where each node is the origin of a submap and the edges in the graph are the coordinate transformations between submaps. Our algorithm attempts to establish additional spatial relationships between submaps corresponding to overlap from parallel tracklines or loop closures.

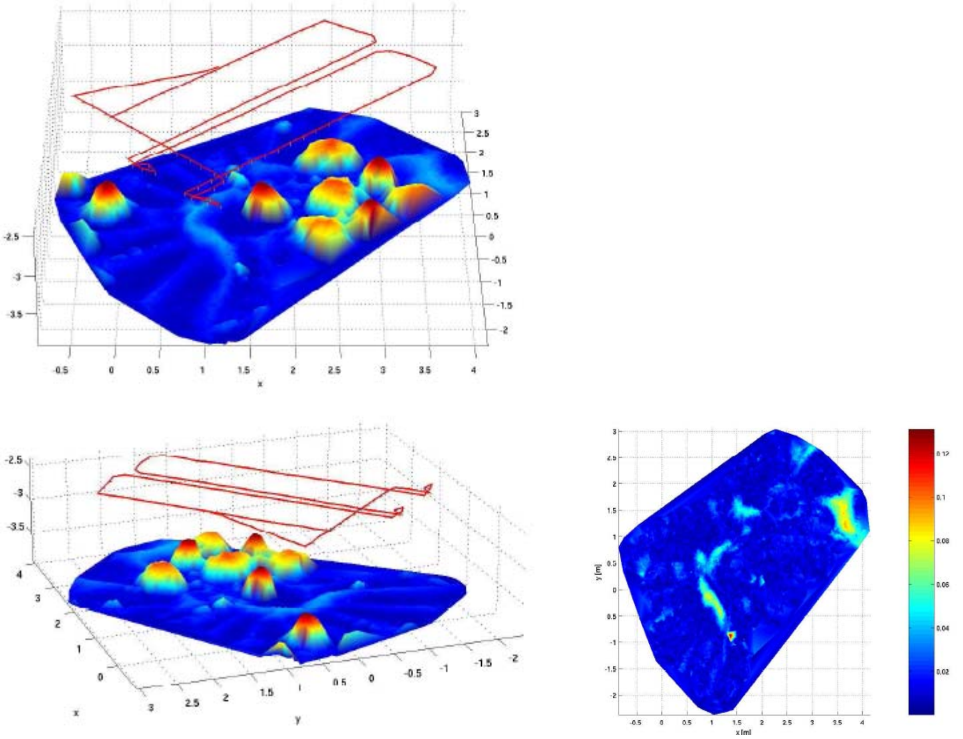

Figure 5. (Left) Two views of the reconstruction of poses and structure for the JHU tank dataset. The camera poses are connected by a red line. A Delaunay triangulation interpolates a surface between 3D feature points. The structure is color-coded according to height. Units are in meters. (Right) Distance map from SFM 3D points to the ground truthed laser scan after ICP registration. Areas of large discrepancies tend to correspond to the carpet being buoyant for the visual survey. An outlier in the reconstruction produced the large error visible at approximate x=1.4m,y=0.8 m.

While the sparse set of 3D points contained in the submaps do not consistently offer discriminating structure, the very fact that they exist as 3D points implies that their appearance in multiple views is characteristic enough to effectively establish correspondences and be reconstructed by the SFM algorithm. We therefore extend the feature description and similarity based matching between images to matching submaps by relying on the appearance of 3D points to propose corresponding features between submaps. The average of the descriptors of the 2D neighborhoods on all views is used as the appearance of the 3D point. The underlying assumption is that a similarity measure which was effective to match 3D points along track will also be effective when matching across submaps. Corresponding 3D points are proposed based on appearance and a robust registration using RANSAC with Horn's algorithm[20] is used to determine which points are in correspondence and the transformation parameters.

The search of additional links continues until no links are left to check or an upper limit is reached (we use eight times the number of submaps for sparse, locally-connected graphs). The submaps are then placed in a global frame by minimizing the descrepancies between composed global estimates and the transformations between submaps. Additional cost terms consider the navigation prior.

Once submaps are in a global frame, camera poses within submaps can also be placed in the global frame. These camera poses are then used to triangulate the location of 3D features. Sparse bundle adjustment then refines both camera poses and 3D feature locations.

Figure 5 illustrates this process. The results are from a survey performed in the Johns Hopkins University (JHU) Hydrodynamics Test Facility using the JHU ROV. As shown in the figure the results were highly consistent with ground-truth obtained by draining the test tank and laser scanning the scene geometry. We have also obtained similar results from a survey using the Seabed AUV at a coral reef off of Bermuda[5].

Self Consistent Bathymetric Mapping

Another application of our techniques arises from the case of multibeam mapping[16] where the areas of interest encompass several square kilometers that are typically mapped with a sonar with ten centimeter sensor accuracy but where the navigation from the standard combine of long baseline transponders and inertial navigation is only good to a meter. To avoid this navigation limitation we break the total mapping problem down into small pieces, each of which contains internal errors typical of the mapping sonar rather than the navigation [19]**.** This is accomplished by assembling small bathymetry *sub-maps* using only the short term dead reckoning information provided by the vehicle navigation sensors. Algorithmically this is accomplished using a delayed state Extended Kalman Filter (EKF) and a simple constant velocity dynamic model of the vehicle motion. This simple model is sufficient given the slow dynamics typical of underwater survey vehicles. The current estimate of the filter state vector contains the position and velocity information required for a 6 degree of freedom (DOF) state estimate.
The delayed portion of the state vector is used to archive historical poses of the vehicle which serve as local 6 DOF origins for the small sub-maps.

After accounting for issues specific to acoustic sensors such as possible errors associated with weak returns, beam patterns effects resulting in the acoustic pulse not striking the bottom, and other false detections, we can approximate the sonar as a three dimensional line scanner. These line scans are assembled into sub-maps using the range data and the vehicle position estimates extracted from the state vector at the time each sonar ping is taken (Figure 6).

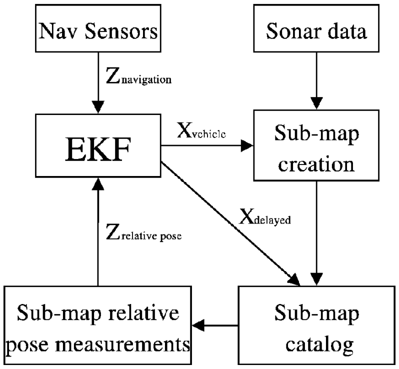

Figure 6. The delayed state EKF block diagram. The sub-mapping algorithm utilizes vehicle navigation data to create small bathymetric sub-maps. The sub-map origins are held in a delayed state vector and used to create relative pose measurements that reduce navigation error.

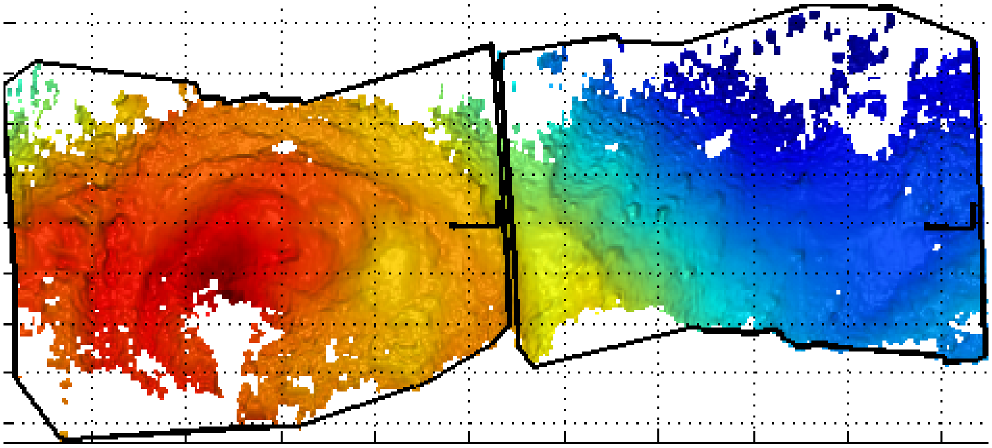

Figure 7. Two sample sub-maps showing their outlines and local reference frames. Note the fine scale features that can be resolved individually within each sub-map. Normal smoothed navigation tends to blur these real and often significant features. Cf Figure 9.

The individual beam ranges are projected and kept as points (soundings) in 3D dot clouds referenced to the delayed state local origins (Figure 7). Sub-maps are sized by collecting pings in this manner over short time scales during which the vehicle position error associated with the inertial navigation is considered small. A sub-map is broken, and a new one started, when the one of several condition are met. A map will be broken when the terrain contained in the map has sufficient 3D structure, the map has become too large or has poor outline shape or, when the estimate of vehicle position uncertainty relative to the map origin becomes greater than a threshold consistent with mapping sonar accuracy. The position error based end condition is designed to keep the navigation errors from corrupting the mapping data.

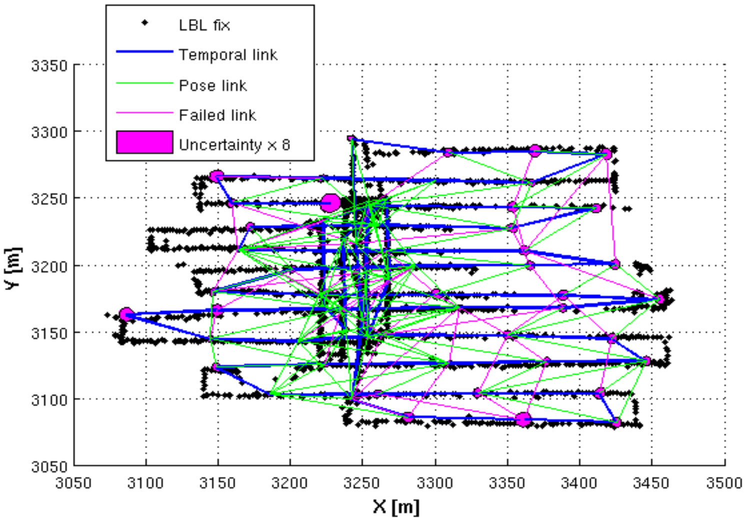

Figure 8. Sub-mapping pose network. This pose network was established by the sub-mapping algorithm. Nodes indicate the location of the sub-map origins. Blue links indicate consecutive poses in time. Green links indicate where relative pose measurements were made. Magenta links indicate links that were tried but not established. The uncertainty ellipses have been scaled in size by 8 times for visibility. Note that the poses fall into alignment with the LBL fix locations even though this algorithm did not utilized LBL measurements. This survey consisted of 62 sub maps and 92 established links.

The mapping error growth associated with navigation errors across the entire survey is reduced by registering the terrain sub-maps to one another and generating relative position measurements between previously visited vehicle states. The sub-maps are compared and aligned using a 2 DOF translation based on correlation followed by a 6 DOF alignment with a modified Iterative Closest Point (ICP) [17][18] approach. Improved registration results are obtained by selecting matching points bases on the quality of the individual sonar returns[19][21][22]. The end result of the algorithm is a constraint network**,** between the sub-map origins (Figure 8). This network enforces consistency based on the sub-map alignments and helps to significantly reduce the inconsistency that would be present if navigation was used alone to produce the terrain map.

Another application of our techniques arises from the case of multibeam mapping[21] where the areas of interest encompass several square kilometers that are typically mapped with a sonar with ten centimeter sensor accuracy but where the navigation from the standard combination of long baseline transponders and inertial navigation is only good to a meter.

The results of our framework are illustrated in Figures 9 and 10 using data collected by the Jason ROV at the TAG Hydrothermal Vent Site located at a depth of 2600 meters on the mid-ocean ridge in the Atlantic Ocean. One can see that the resulting map is a far better representation of the environment. We have also used consistency within the submaps to derive corrected navigation estimates for the vehicle trajectory over the course of the survey.

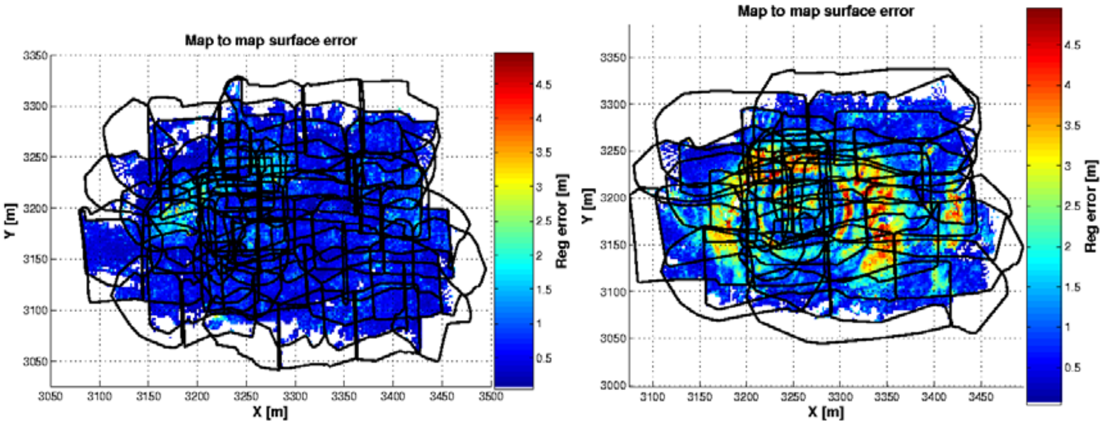

Figure 9. Error in bathymetric mapping as measured by self consistency across multiple sub-maps. (Left) Map to map surface error for our algorithm versus (right) map to map surface error using a standard smoothing method.

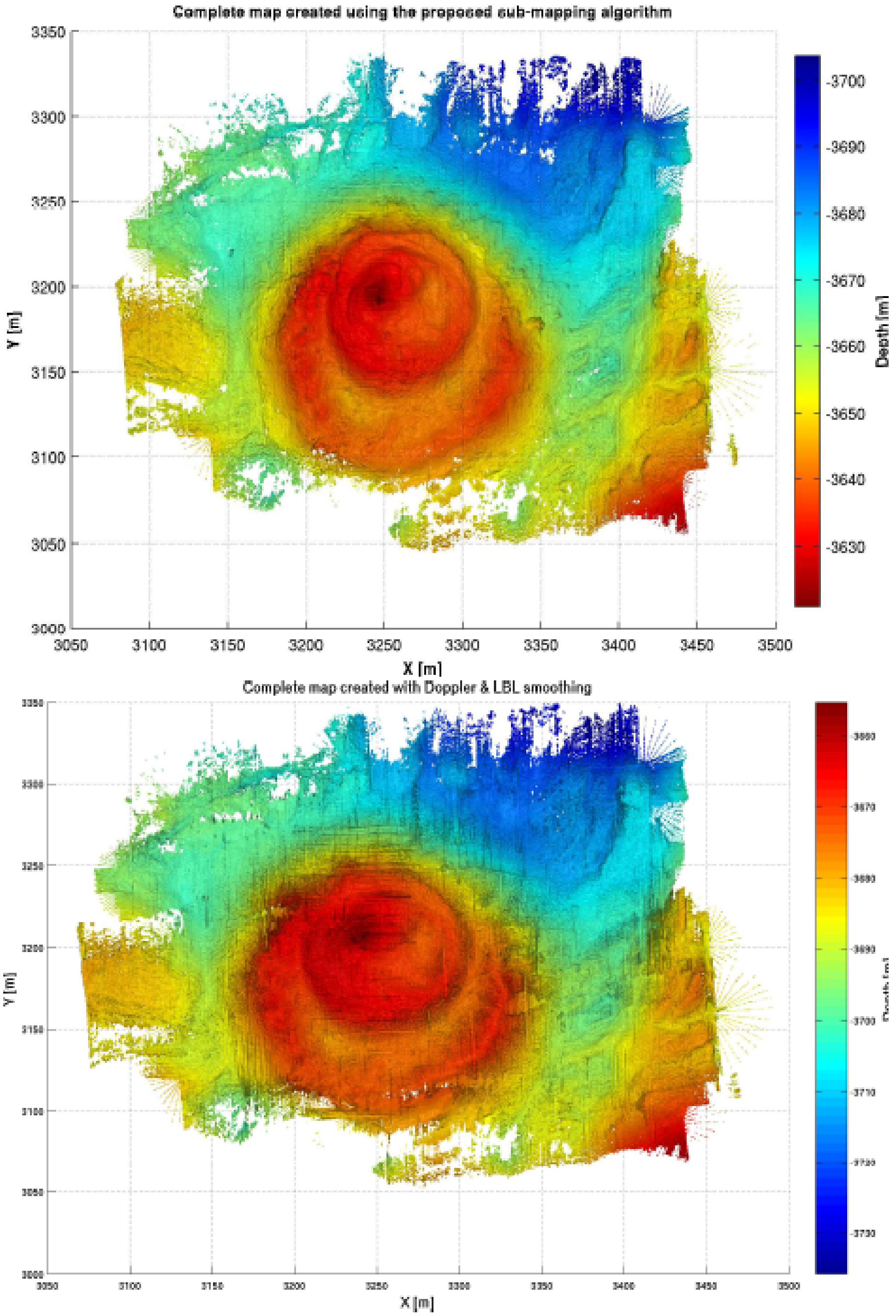

Figure 10. A comparison between the terrain map created using sub-mapping, *(top)* , and a version of the map created using a standard smoothing method *(bottom).* The sub-map created map shows significantly less registration error and sonar scan patterning. The sub-mapped version also brings out details that were lost in the smoothed map due to misregistration.

Visually Augmented Navigation

We can further build upon the delayed-state EKF framework and two-view structure from motion case to formulate a vision-based simultaneous localization and mapping (SLAM) approach to providing high precision, accurate navigation measurements on an underwater robotic vehicle. Similar to the bathymetry-based EKF submapping strategy, our methodology is to employ pairwise-generated image measurements as spatial constraints in a graph over a collection of historical vehicle poses. However, because we are able to generate camera measurements at the pairwise level, we choose to instead maintain all pose samples that are associated with image acquistion (Figure 11). This differs from the aggregate submapping strategy used for bathymetry-based navigation and implies that the EKF's scalability becomes a severe issue (due to the quadratic complexity of maintaining the covariance) as the image-based navigation uses orders of magnitude more delayed-states.

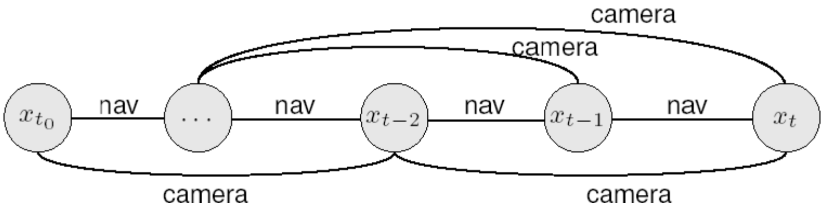

Figure 11. A conceptual depiction of the delayed-state graph network and its constraints.

A well known and very attractive property of formulating SLAM in the information form is that the information matrix (as in Fisher information) has the direct interpretation as a Gaussian graphical model[23][24]. Sparsity in this model (i.e., missing edges) implies available conditional independencies in the joint-distribution, which can be exploited to realize efficient inference. While others have shown that the feature-based SLAM information matrix obeys a "close-to-sparse" structure when properly normalized [14][25] in our formulation of view-based SLAM[15], the information matrix is exactly sparse without having to make any sparse approximations. This implies that for a bounded graph structure, as is the case with typical underwater surveys, view-based SLAM systems comprise a sparse information parameterization without incurring any sparse approximation error.

Based upon this insight, we have implemented a view-based SLAM system for underwater applications built around fusing 6-DOF relative-pose camera measurements from monocular overlapping sea floor imagery with traditional underwater vehicle dead-reckon navigation sensors. Our state vector consists of samples from the robot's trajectory acquired at image acquisition and is maintained using a sparse extended information filter. We use our two view image registration engine to provide non-Markov edge constraints in the corresponding pose network. These "spatial" edges constrain the pose graph and enforce global consistency from local constraints. This system was tested with data collected using the *Hercules* ROV operating at a depth of 3750 meters at the wreck of the *RMS Titanic*. The survey covered an area of about 3100 square meters on the seafloor with a accumulated survey path length over 3.4 kilometers. Results are shown in Figure 12.

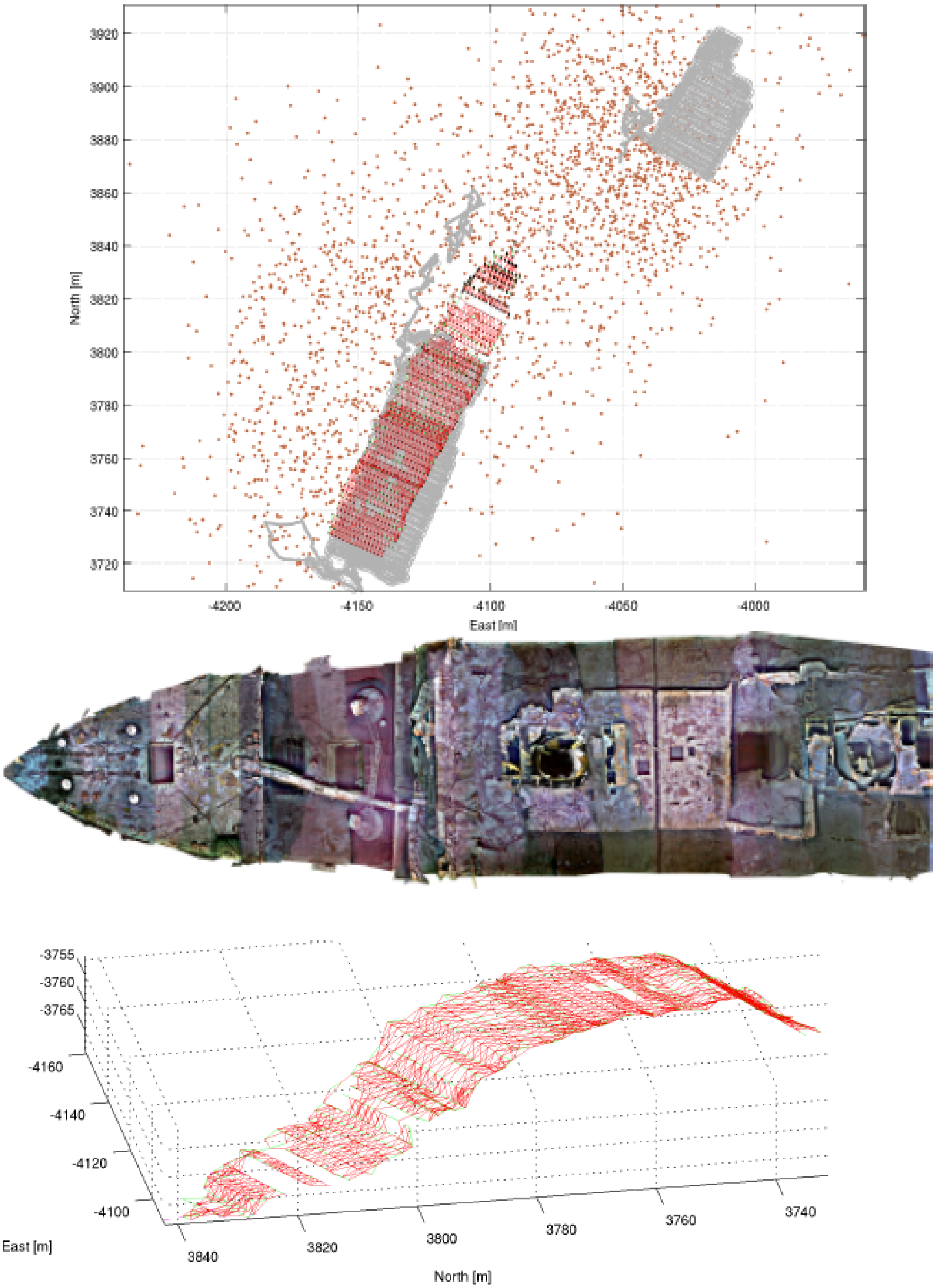

Figure 12 Mapping the RMS Titanic. *(Top)* An XY plot comparing the raw dead-reckon navigation data (gray), shipboard ultra-short baseline tracking (brown), and reconstructed survey trajectory from a vision-based 6 DOF SLAM information filter (red). *(Middle)* A photomosaic of the RMS Titanic constructed from the digital still images and *(Bottom)* the complete 6 DOF visually based navigation results for the entire survey.

Conclusions

In this paper we have highlighted some of the fundamental issues associated with the lack of precise and accurate navigation and how they affect our ability to conduct high resolution mapping efforts in the deep sea. We have presented three different applications of systems level approaches for deep water mapping that exploit vehicle attitude and navigation information and enforce local and global consistency within sensor measurements to yield superior mapping results commensurate with sensor accuracy. While improving mapping fidelity these methods also provide us with independent mechanisms for ground truthing, refining and bounding the coarse navigation estimates that are typical in the deep ocean.

These algorithms are applicable across the entire suite of imaging and robotic underwater vehicles - manned, towed, tethered and autonomous. Our work in these areas is continuing with an emphasis on implementing a number of these algorithms in real-time on board the vehicles to better help us exploit our precision mapping algorithms for real-time adaptive surveys.

Acknowledgements

The authors gratefully acknowledge our multitude of collaborators on the following research expeditions from which data for this paper was gathered – Dr Rob Reves Sohn (Bermuda Coral and TAG), Prof Louis Whitcomb (JHU Hydrodynamics Facility), and Prof Robert Ballard (Titanic). This work was funded in part by a grant from the National Science Foundation’s Censsis ERC under grant # EEC-9986821.

References

[1] D.R. Yoerger, A.M. Bradley, M.-H. Cormier, W.B.F. Ryan, and B.B.Walden. Fine-scale seafloor survey in rugged deep-ocean terrain with an autonomous robot. In *IEEE International Conference on Robotics and Automation*, volume 2, pages 1767–1774, San Francisco, USA, 2000.

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

[3] J. Howland. Digital Data Logging and Processing, Derbyshire Survey, 1997. Technical report, Woods Hole Oceanographic Institution, December 1999.

[4] National Transportation Safety Board, Washington, DC. *Aircraft Accident Brief: EgyptAir Flight 990, Boeing 767-366ER, SU-GAP, 60 Miles South of Nantucket, Massachuesetts, October 31, 1999*, 2002. Aircraft Accident Brief NTSB/AAB-02/01.

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

[6] P. H. Milne, *Underwater Acoustic Positioning Systems*. Gulf Publishing Company, Houston, 1983.

[7] L. Whitcomb, D. Yoerger, and H. Singh, .Advances in Doppler-Based Navigation of Underwater Robotic Vehicles,. in *Proceedings of the 1999 International Conference on Robotics and Automation*, vol. 1, 1999, pp. 399.406.

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

[9] M. Pollefeys, R. Koch, M. Vergauwen, and L. Van Gool. Hand-held acquisition of 3d models with a video camera. In *Second InternationalConference on 3-D Digital Imaging and Modeling*, pages 14–23, Los Alamitos, CA, 1999. IEEE Computer Society Press.

[10] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, “Bundle Adjustment – a Modern Synthesis,” in Vision Algorithms: Theory and Practice, ser. Lncs, W. Triggs, A. Zisserman, and R. Szeliski, Eds. Springer-Verlag, pp. 298–375. 2000

[11] A.W. Fitzgibbon and A. Zisserman, Automatic Camera Recovery for Closed or Open Image Sequences, *Proceedings of the 5th EuropeanConference on Computer Vision*, pages 311–326, Freiburg, Germany, June 1998.

[12] C.C. Slama, editor. *Manual of Photogrammetry*. American Society of Photogrammetry, Bethesda, MD, fourth edition, 1980.

[13] O. Pizarro, R. Eustice, and H. Singh, “Relative Pose Estimation for Instrumented, Calibrated Imaging Platforms,” in Proceedings of Digital Image Computing Techniques and Applications, Sydney, Australia, pp. 601–612, December 2003.

[14] S. Thrun, Y. Liu, Z. Koller, H. Ghahramani, H. Durrant-Whyte, and A. G. Ng, .Simultaneous mapping and localization with sparse extended information filters,. *International Journal of Robotics Research*, vol. 23, pp. 693.716, 2004.

[15] R. Eustice, H. Singh, and J. Leonard, .Exactly Sparse Delayed-State Filters,. in *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, Barcelona, SPAIN, April 2005.

[16] H. Singh, L. Whitcomb, D. Yoerger, and O. Pizarro, .Microbathymetric Mapping from Underwater Vehicles in the Deep Ocean,. *Computer Vision and Image Understanding*, vol. 79, no.1, pp.143.161, 2000.

[17] P. J. Besl and N. D. McKay, A method for registration of 3-d shapes, *IEEE PAMI*, vol. 14, pp. 239-256, 1992.

[18] S. Rusinkiewic and M. Levoy, Efficient variants of the icp algorithm, in *Proceedings IEEE 3DIM*, Canada, 2001, pp. 145.152.

[19] C. Roman, Self Consistent Bathymetric Mapping from Robotic Vehicles in the Deep Ocean, Ph.D. Thesis, MIT/WHOI Joint Program in Oceanographic Engineering, May 2005.

[20] B. Horn, .Closed-Form Solution of Absolute Orientation Using Unit Quaternions,. *Journal of the Optical Society of America*, vol. 4, no. 4, pp. 629.642, April 1987.

[21] C. Roman, H. Singh, Improved vehicle based multibeam bathymetry using sub-mapping and SLAM, *IEEE Intelligent Robotics and Systems (IROS)*, 2005 *to appear*.

[22] H. Medwin, H., C. Clay, *Fundamentals of Acoustical Oceanography,* Academic Press, London, 1998.

[23] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufman Publishers, San Mateo, CA. 1988.

[24] Y.Weiss, W.T. Freeman, Correctness of belief propagation in Gaussian Graphical Models of Arbitrary Topology. *Neural Computation*, 13(10):2173-2200, 2001.

[25] U. Frese, A Proof for the Approximate Sparsity of SLAM Information Matrices, Proceedings of IEEE International Conference on Robotics and Automation, pp 331-337, Barcelona, Spain, 2005.