Large area 3D reconstructions from underwater surveys

Oscar Pizarro, Ryan Eustice and Hanumant Singh
Deep Submergence Laboratory
Woods Hole Oceanographic Institution
Woods Hole, MA 02543
Email: opizarro@whoi.edu

***Abstract*— Robotic underwater vehicles can perform vast optical surveys of the ocean floor. Scientist value these surveys since optical images offer high levels of information and are easily interpreted by humans. Unfortunately the coverage of a single image is limited by absorption and backscatter. There is a need to present an overall view of the survey area. Recent work on underwater mosaics assume planar scenes and are applicable only to situations without much relief.**

We present a complete and validated system for processing optical images acquired from an underwater robotic vehicle to form a 3D reconstruction of the ocean floor. Our approach is designed for the most general conditions of wide-baseline imagery (low overlap and presence of significant 3D structure) and scales to hundreds of images. We only assume a calibrated camera system and a vehicle with uncertain and possibly drifting pose information (from, for example, a compass, depth sensor and a Doppler velocity log).

Our approach is based in a combination of techniques from computer vision, photogrammetry and robotics. We use a local to global approach to structure from motion, aided by the navigation sensors on the vehicle to generate 3D submaps. These submaps are then placed in a common reference frame that is refined by matching overlapping submaps. The final stage of processing is a bundle adjustment that provides the 3D structure, camera poses and uncertainty estimates in a consistent reference frame.

We present results with ground-truth for structure as well as results from an oceanographic survey over a coral reef covering an area of approximately one hundred square meters.

I. Introduction

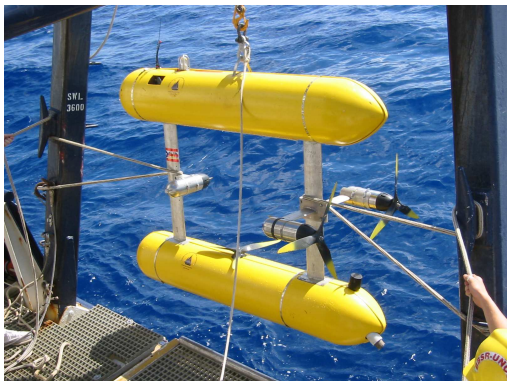
A. Context

Optical imaging of the ocean floor offers scientists 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 only a few square meters. For many scientific surveys, however, the area of interest is large, and can only be covered by hundreds or thousands of images acquired from a robotic vehicle or towed sled. Such surveys are required to study hydrothermal vents and spreading ridges in geology [1], ancient shipwrecks and settlements in archeology [2], forensic studies of modern shipwrecks and airplane accidents [3] [4], and surveys of benthic ecosystems and species in biology [5] [6].

The visible spectrum in water has attenuation lengths of the order meters, thus most underwater vehicles carry out optical imaging surveys using their own light source. Apart from casting shadows that move across the scene as the vehicle moves, power and/or size limitations lead to lighting patterns that are far from uniform. Also with the advent of autonomous underwater vehicles (AUVs) for imaging surveys [1] [6] additional constraints are imposed by their limited energy budgets. AUV surveys are typically performed with strobed light sources rather than continuous lighting, and acquire low overlap imagery in order to preserve power and cover greater distances.

Generating a composite view by exploiting the redundancy in multiple overlapping images is usually the most practical and flexible way around this limitation. Recent years have seen significant advances in mosaicing [7] [8] and full 3D reconstruction [9] [10] [11] though most of these results are land based and do not address issues particular to underwater imaging. Underwater mosacing has been motivated largely by vision-based navigation and station keeping close to the seafloor [12] [13] [14]. The large-area mosaicing problem with low overlap under the assumption of planarity is addressed in [15]. Mosaicing assumes that images come from an ideal camera (with compensated lens distortion) and that the scene is planar [16]. Under these assumptions the camera motion will not induce parallax; 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 obvious 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 [17]. We propose that when dealing with a translating camera over non-planar surfaces, recovering 3D structure is the proper approach to providing a composite global view of an area of interest. The same challenges seen in mosaicing underwater apply to SFM underwater with the added requirement that scene points must be imaged at least twice to produce a roughly uniform distribution of reconstructed feature points through triangulation (50% overlap in the temporal image sequence). These techniques are considerably more complex than mosaicing: even for



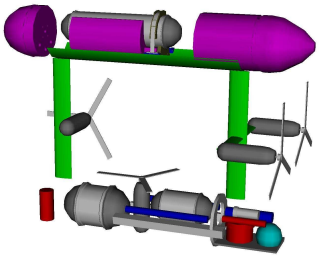


Fig. 1. The Seabed vehicle in the Bermuda 2002 cruise. CAD views showing the vehicle without shells.

land-based applications (with high overlap, structured motion and uniform lighting) consistency at large scales can not be guaranteed unless other sensors are available. Some promising work has gone into 3D image reconstruction underwater [18] using a stereo-rig with high overlap imagery in a controlled environment.

Underwater vehicles for scientific surveys use navigation sensors that provide pose estimates. This information can be used to constrain and regularize the underwater structure from motion problem. In previous work [19] [20] we show in detail how to improve the search for corresponding features between images. In addition, we use navigation sensors to provide estimates of baseline magnitude and to select a unique solution in cases where imagery provides multiple valid solutions.

B. Imaging Platform

The Seabed AUV acquired the field data used in this thesis (Figure 1). The vehicle was designed as a calibrated and pose-instrumented platform for underwater imaging. Seabed is capable of maneuvering at slow speed and passively stable in pitch and roll. The vehicles specifications are summarized in Table I. Seabed collected the field data used in this paper following survey patterns preprogrammed as a mission and executed in dead-reckoning mode. The vehicle makes acoustic measurements of both velocity and altitude relative to the bottom. Absolute orientation is measured within a few degrees using a magneto-inductive compass and inclinometers, while depth is obtained from a preseeure sensor.

Vehicle	
Depth rating	2000 meters
Size	2.0 m (L) $\times$ 1.5 m (H) $\times$ 1.5 m (W)
Mass	200 kg
Maximum Speed	1.2 m/s
Batteries	2 kWh Li-ion pack
Propulsion	four 150 W brushless DC thrusters
Navigation	
Attitude+Heading	Tilt $\pm 0.5^{\circ}$, Compass $\pm 2^{\circ}$
Depth	Paroscientifi c pressure sensor, 0.01%
Velocity	RDI Navigator ADCP $\pm 1 - 2mm/s$
Angular rates	Crossbow 3-axis gyro
Altitude	RDI Navigator
Optical Im.	
Camera	Pixelfly 12bit 1280$\times$1024 CCD
Lighting	one 200 Ws strobe
Separation	1m between camera and light
Acoustic Im.	
Sidescan sonar	MST 300 kHz (300 m depth rating)
Pencilbeam sonar	Imagenex 881 675 kHz
Other Sensors	
CTD	Seabird 37SBI

TABLE I
SUMMARY OF THE SEABED AUV SPECIFICATIONS.

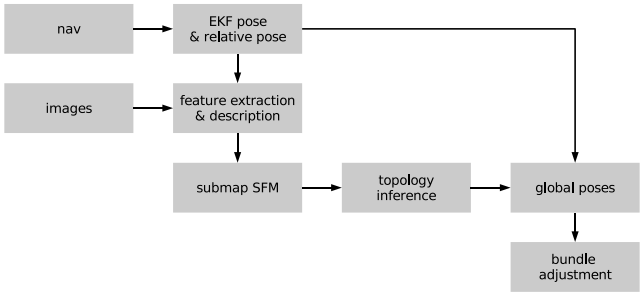


Fig. 2. Flowchart of structure and motion recovery from underwater imagery. An image sequence is processed into short submaps of structure and motion aided by navigation information. Submaps are then matched to infer and refi ne additional spatial constraints (such as loop closures and parallel tracklines). An initial guess of poses and structure in a global frame is then used to perform a fi nal bundle adjustment.

C. Outline

Our methodology (Figure 2) takes a local-to-global approach inspired by mosaicing [21] and structure from motion (SFM) [11] [22] but takes advantage of navigation and attitude information. Local subsequences are derived independently and then registered in a global frame for bundle adjustment. Our approach seems more suitable than pure sequential methods [23] because in an underwater survey each 3D feature appears only in a few images making the global solution more like a series of weakly correlated local solutions.

The following section briefly describes our approach focusing on feature extraction and description, robust two view relative pose estimation, submap generation, topology exploration and local to global registration. The last section presents results from a coral reef survey and validation of the proposed framework by tank experiments with ground truth.

II. Algorthim Description

Our description follows the block diagram presented in Figure **??**.

A. Feature Extraction and Description

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 [24] 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 using a modified version of the method proposed by Tuytelaars [25]. 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 (Figures 3 and 4).

Features are then represented compactly using moment-based descriptors [26], which have shown promise in describing image regions for matching purposes. We chose to use Zernike moments as descriptors as they are compact (generated from an orthogonal complex polynomials) and highly discriminating [27] [15]. Typical applications only use the magnitude of Zernike moments as this provides rotational invariance, but we can pre-compensate for orientation using attitude sensors and therefore utilize the full complex moments.

For feature matching we derive the proper weighting of the Zernike moments such that the dot product of the vector of weighted moments approximates the correlation score for the original patches (warped into a disc) [28].

Tests with real data demonstrate that the affine-invariant features offer improved matching under wider viewing angles (Figure 5).

B. Submap Generation

The core of the algorithm for SFM is based on robust estimation of the essential matrix (Figure 6) [19]. 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 (Figure 7) by propagating pose and altitude uncertainties through the two view point-transfer equation [20].

A modified version of RANSAC [29] determines the correspondences which are consistent with that essential matrix and the essential matrix consistent with the inliers (Figure **??**). In cases of multiple valid solutions the solution closest to the navigation-based prior in the Mahalanobis sense is selected. 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 9).

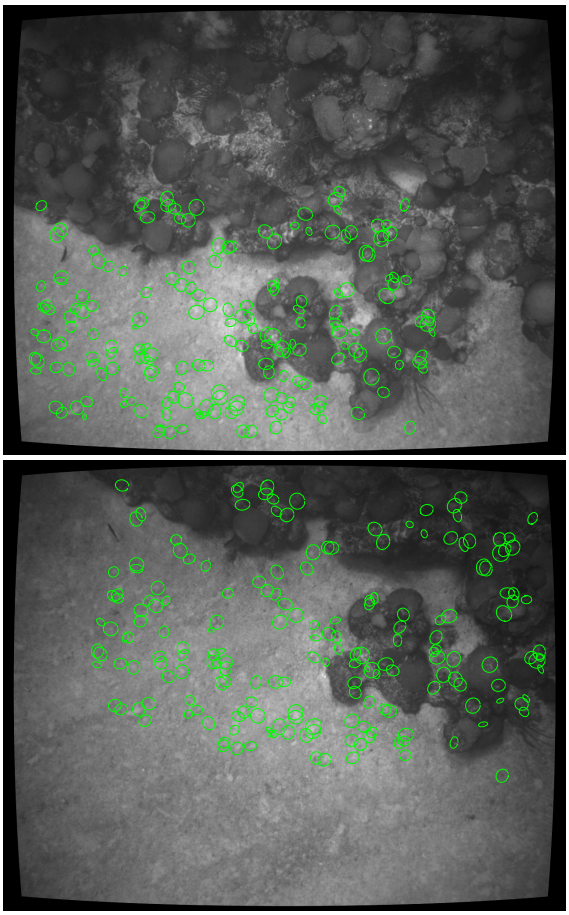


Fig. 3. Affine invariant regions extracted using a modified version of the method proposed by Tuytelaars. Only regions that are found in correspondence are shown.

In cases where scene elements are viewed in three (or more) views the algorithm attempts to obtain the pose of the third view by robust resection [29] (Figure 10), otherwise the two view essential matrix estimation is used. The submap is generated by incorporating each image in the temporal sequence by resection or two view estimation until a maximum number of 3D features have been instantiated. The submap is then closed and bundle adjusted. We use submaps with 1500 to 2000 3D features as a compromise between the complexity of individual submap bundle adjustment and the complexity of the network of submaps formed (and the number of submap matching operations).

C. Global Representation

The temporal sequence of images is processed into a set of 3D submaps with estimates of coordinate transformations between temporally adjacent submaps. 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 (Figure 11). Our algorithm attempts to establish ad-

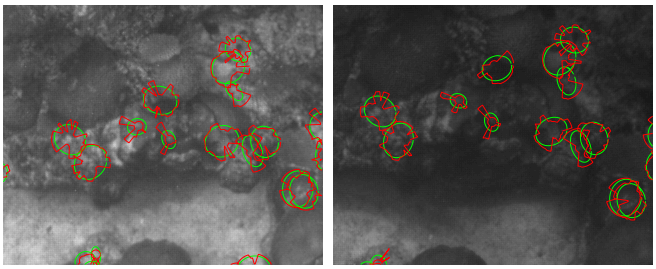


Fig. 4. Detail of some of the extracted regions in figure 3. The actual border samples are connected with red lines. The elliptical region that approximates the border samples is shown in green.

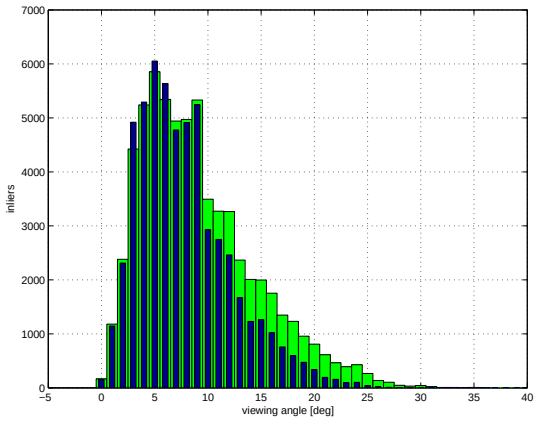


Fig. 5. For the matches considered inliers it is possible to calculate the viewing angle change between each camera to the feature. For all matches, across all pairs in the trial (over 350), we show the number of inliers as a function of viewing angle. For narrow-baseline conditions (angles of 10° or less) both regions behave similarly. For larger viewing angles the affine invariant region (green) outperforms the fixed window (blue).

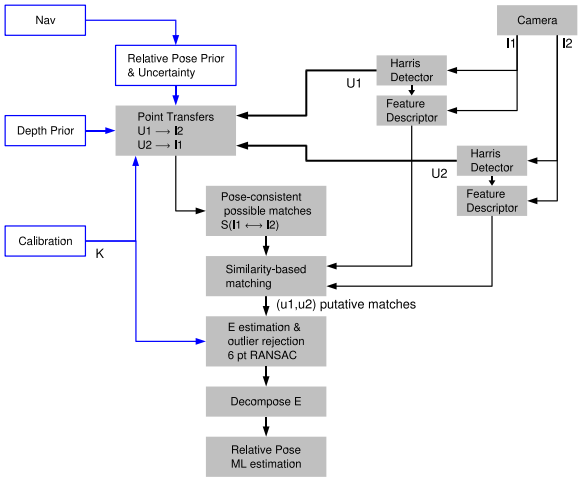


Fig. 6. Overview of our approach to relative pose estimation from instrumented and calibrated platforms. Unshaded blocks represent additional information compared to the uninstrumented/uncalibrated case. Given two images, we detect features using the Harris interest point detector. For each feature we then determine search regions in the other image by using sensor based pose and depth information. Putative matches are proposed based on similarity and constrained by regions. We then use RANSAC and the proposed 6-point algorithm to robustly estimate the essential matrix which is then decomposed into its proper motion parameters. The pose is then refined by minimizing the reprojection error over all matches considered inliers.

ditional spatial relationships between submaps (corresponding to overlap from parallel tracklines or loop closures). This is performed by placing submaps in a common reference frame by composing transformations along paths of minimum uncertainty using Dijkstra's algorithm [30].

Submaps must be matched in order to establish new edges in the graph. Registering two sets of 3D points with unknown correspondences is traditionally performed with Iterative Closest Point (ICP) techniques [31]. In its strictest sense, ICP is only a refinement of the transformation between two sets of 3D points that are already relatively well aligned and in which all points in one set have a match in the other.

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 (Figure 12) 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 [32] is used to determine which points are in correspondence and the transformation parameters (Figure 13).

The search of additional links continues until no links are left to check or an upper limit is reached (typically $8N$). 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 subamps 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 [33] [9] then refines both camera poses and 3D feature locations.

To illustrate this process we present in Figure 14 the resulting structure from a survey performed in the Johns Hopkins University (JHU) Test tank. The tank had a carpet drapped over the bottom and real and artificial rocks of varying size placed on the bottom to simulate an underwater scene with considerable 3D structure. The evolution of the submap graph for that reconstruction is conveyed in Figure 15 while the reprojection errors for the structure is presented in Figure 16.

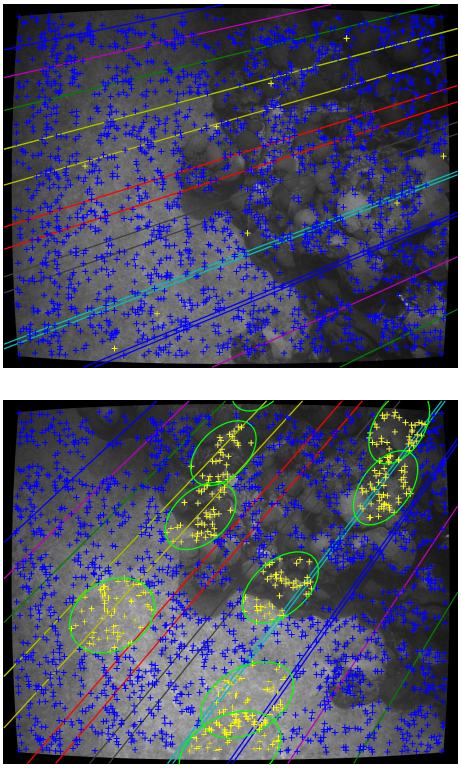


Fig. 7. Prior pose restricted correspondence search on a pair of underwater coral reef images. (left) Interest points are shown in blue. A sampling of interest points (yellow) is transfered to the right image. (right) The 99% confi dence regions for the transfered points based on the pose prior and depth standard deviation of 0.75m. The candidate interests points which fall within these regions are highlighted in yellow.

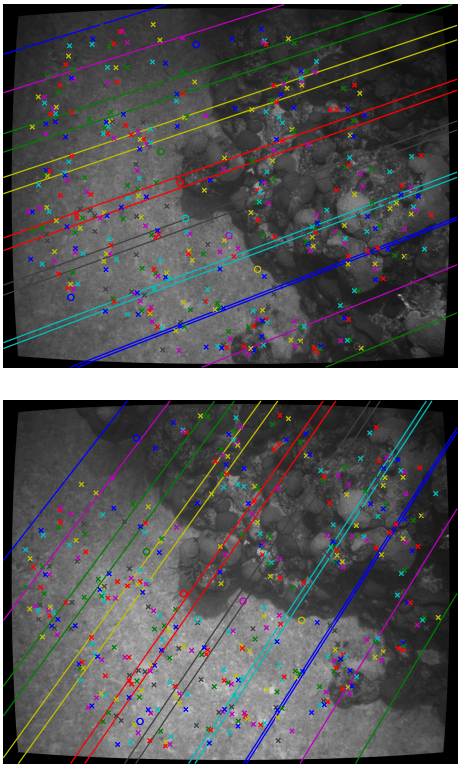


Fig. 8. Epipolar geometry and correspondences. The given image pair illustrates the MAP refi ned image-based epipolar geometry. RANSAC determined 398 consistent inliers designated 'x', from the putative set of 405 matches. The rejected outliers are designated 'o'.

III. Validation and Results

A. JHU Tank Structure ground truth

For validation purposes the tank used in Figure 14 was drained and scanned with a Leica Geosystems - HDS2500 (serial number P24) laser scanner. The registered model of the tank has more than 3.8 million points with an estimated accuracy of 1.2 mm. The surface area was approximately 41m^2 resulting, on average on 9 range measurements for each cm^2 of the bottom.

We initially aligned SFM reconstruction with the laser data by selecting easily recognizable landmarks (Figure 17) and then refined through ICP. The carpet was slightly buoyant underwater and was kept on the bottom by multiple lead weights and that after the tank was drained the carpet settled under its own weight. We attempted two registration strategies to overcome the non-rigid transformation between surfaces: using only points belonging to rocks to register (segmenting by height under the assumption that the rocks in the scene did not move), and performing ICP based on the points with registration errors below the median error (under the assumption that at least half the points remained fixed). Results were very similar for both strategies and we present the median-based approach since it highlights regions where the

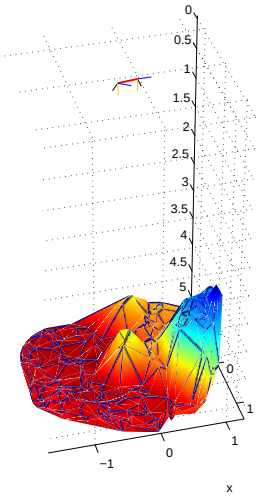
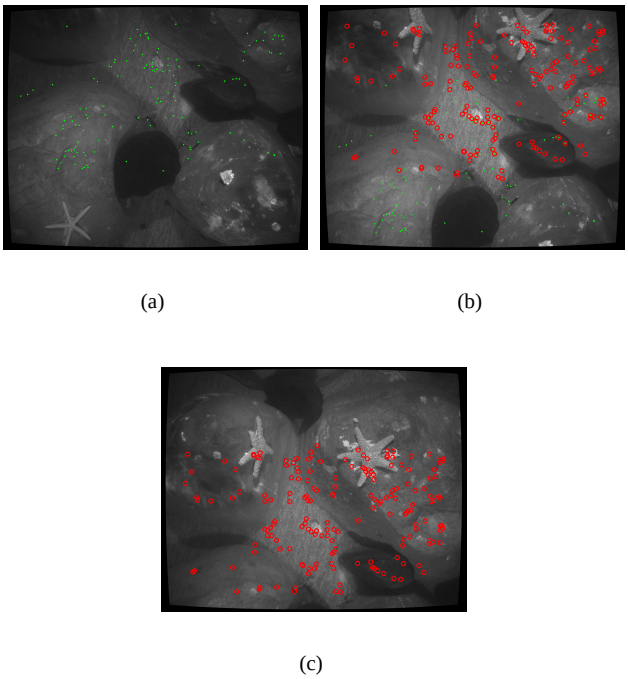


Fig. 9. Triangulated inliers for the pair in fi gure 8. Coordinates in meters, in the reference frame of the fi rst camera.

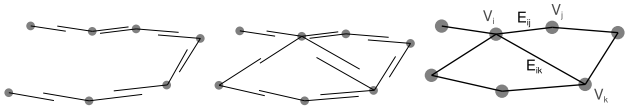


(a) (b)

(c)

(d) (e) (f)

Fig. 10. Illustration of growth of a submap based on resection. Images (a) and (b) have corresponding features marked by green dots. The structure and motion implied by those correspondences is illustrated in (d) with units in meters. Images (b) and (c) have correspondences marked by red circles. The features viewed by the three images are marked by both a green dot and a concentric red circle. (e) These features are used in resection to initialize the pose of the third camera. (f) Then the additional correspondences between (b) and (c) are triangulated and the poses refined.



(a) (b) (c)

Fig. 11. Placing nodes (Gray circles) in a globally consistent frame. From relative transformations (black links) in a temporal sequence (a), to proposing and verifying new additional links (b) to a network with nodes consistent with the relative transformations (c).

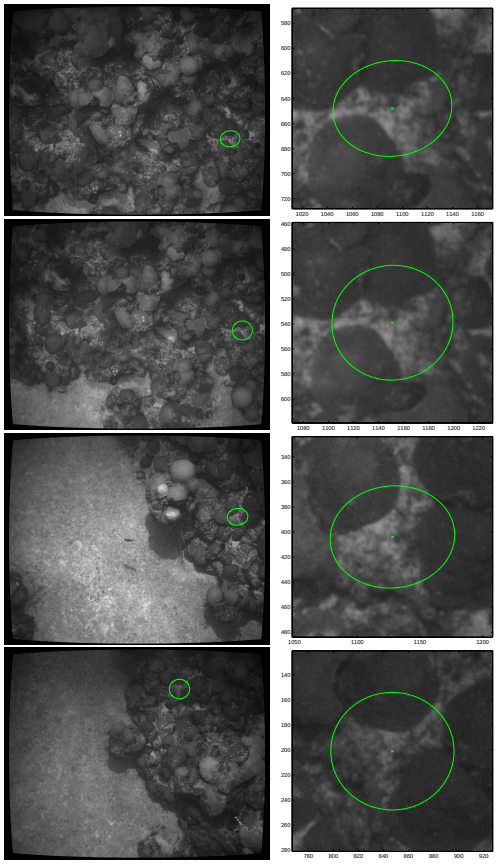


Fig. 12. Multiple views of a 3D feature: the image and the feature neighborhood (left, extracted as described in §**??**) and a detail of around the feature point (right column). The top two rows correspond to images that belong to a submap on the first trackline of the survey, the bottom two rows are from a submap from the second trackline.

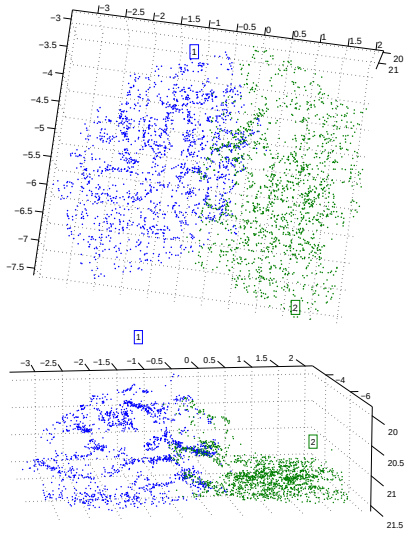


Fig. 13. Views of the registered submaps that contain the images in figure 12. The blue dots (left half) correspond to the 3D features of the submap on the first trackline of the survey. The green dots (right half) correspond to features in a submap on the second trackline of the survey.

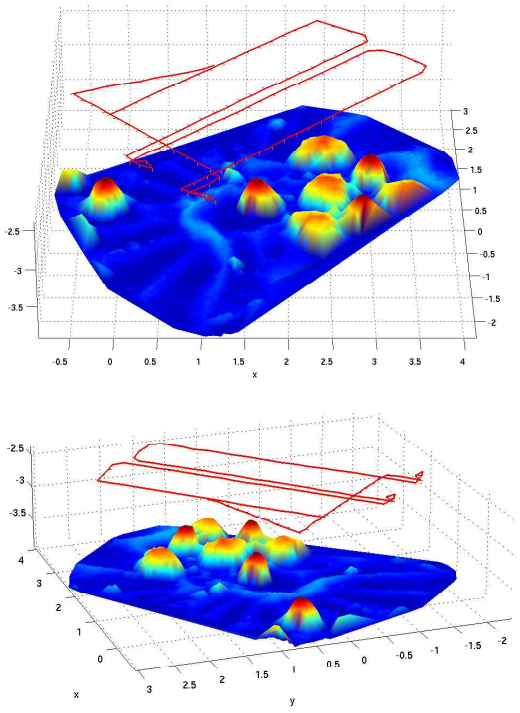


Fig. 14. Two views of the reconstruction of poses and structure for the JHU tank. The camera poses are connected by a red line. A Delaunay triangulation interpolates a surface between 3D feature points. The structure is color-coded according to height. Units are in meters.

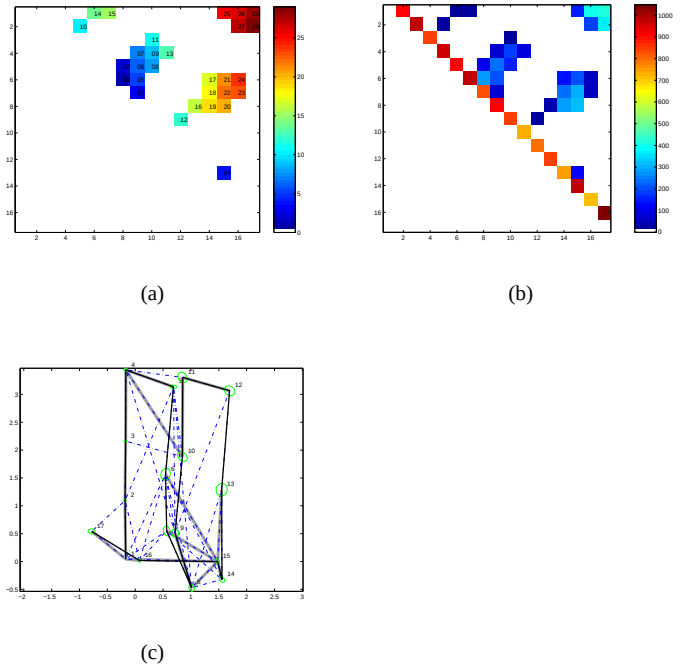


Fig. 15. (a) Order in which links across track were added to the graph. The 'zipper' effect in parallel tracklines is apparent as links close in time are established before more distant ones. (b) The number of matching features between submaps. The closing of the loop can be seen in the relatively high number of common features between the first and last submaps. (c) The plane view of the submap origins according to the shortest path algorithm: the temporal sequence (fine black), the additional links (dot-dashed blue) and the shortest uncertainty path from the origin node (wide gray).

carpet moved.

Figures 18 and 19 indicate that the registration errors are of the order of centimeter level with a 2% change in scale. Though the tank is a relatively small scale reconstruction problem, these results suggest that the approach is capable of delivering reasonable estimates of scene structure.

By using points below the median error to calculate the similarity transformation to register the SFM and laser data we effectively segment the data into two halves, one of which was allowed to deform while the other was not. It is interesting to note from Figure 20 that most of the outliers correspond to the broad carpet waves.

B. Bermuda survey

In August 2002 the Seabed AUV performed several transects on the Bermuda shelf as well as some shallow water engineering trials. This section presents results from a shallow water (20 m approx) area survey programmed with several parallel tracklines for a total path length of approximately 200 m and intending to cover 200 m^2. Due to very strong swell and compass bias the actual path deviated significantly from the assumed path. This data set illustrates the capabilities to infer links in the graph of submaps to yield a consistent reconstruction.

A section of 169 images demonstrates matching and reconstruction along the temporal sequence and across track with multiple passes over the same area. Figure 21 presents Delaunay triangulated surfaces trough the reconstructed points and the camera trajectory. Plan views of the camera trajectory, the links (common 3D features) between views and the uncertainty in the xy position of the cameras are shown in figure 22.

Figure 23 shows features points and the convex hull of the submaps. Spatial overlap between temporally adjacent submaps is consistent while across track overlap is a function of the trajectory followed by the vehicle.

IV. Concluding Remarks and Future Work

We have presented a brief overview of a underwater structure from motion algorithm that takes advantage of vehicle navigation estimates to constrain the image-based solution. The assumed imaging configuration is quite restrictive and makes the image matching problem particularly challenging. This work will be extended to provide dense 3D reconstructions of the ocean floor, which in turn can lead to improved imagery by range-based compensation of absorption.

References

[1] D.R. Yoerger, A.M. Bradley, M.-H. Cormier, W.B.F. Ryan, and B.B. Walden. Fine-scale seafbor 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.

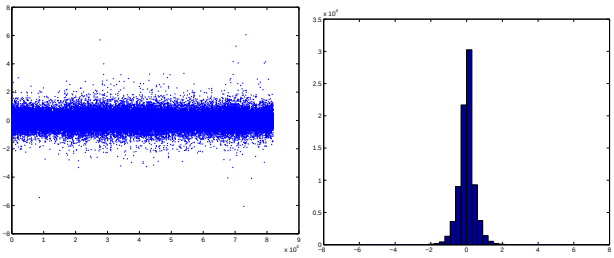


Fig. 16. (Left) The reprojection errors (both x and y coordinates) for all reconstructed features. Some outliers are present though their effect is reduced by using an m estimator in the bundle adjustment.(Right) A histogram of the same errors. For visualization purposes 95% of the features with lowest associated reprojection errors are displayed in the reconstructions of Figure 14.

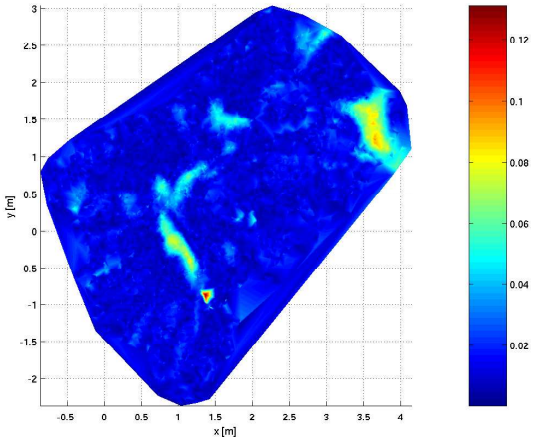


Fig. 18. Distance map from SFM 3D points to the laser scan after ICP registration. Areas of large discrepancies tend to correspond to the carpet being buoyant for the visual survey. An outlier in the reconstruction produced the large error visible at approximate x=1.4 m,y=0.8 m.

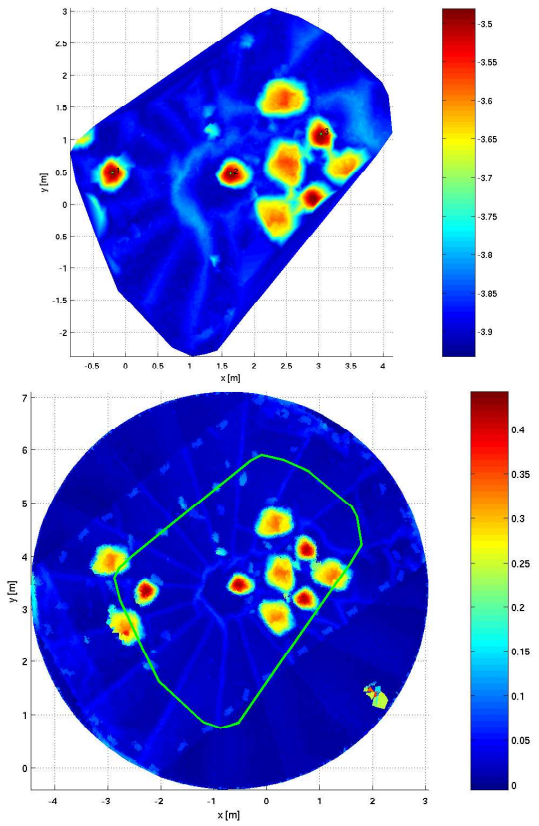


Fig. 17. (Top) Height map from the SFM reconstruction. Surface based on a Delaunay triangulation. The labeled points were manually selected for the initial alignment with the laser scan. (Bottom) Height map from the laser scan. The outline of the manually registered SFM reconstruction is shown in green.

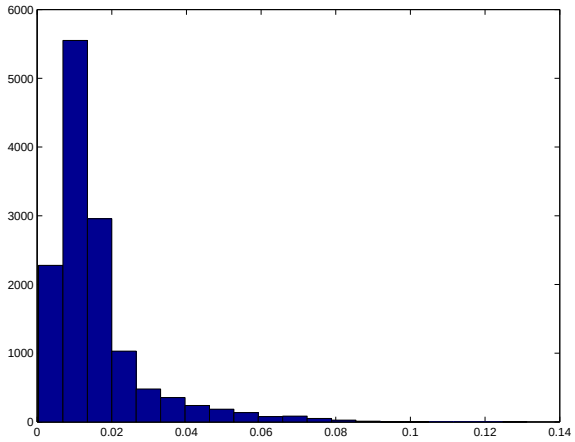


Fig. 19. The distribution of minimum distances to the laser scan from each recovered 3D point. Because of the moving carpet only the points below the median error where used to calculate the registration transformation. The similarity based registration results in an RMS distance of 3.6 cm. And the scale is recovered to within 2%.

[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] C.R. Smith. Whale falls: Chemosynthesis at the deep-sea fbor. *Oceanus*, 35(3):74–78, 1992.

[6] 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.

[7] H.S. Sawhney and R. Kumar. True multi-image alignment and its application to mosaicing and lens distortion correction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(3):235–243, March

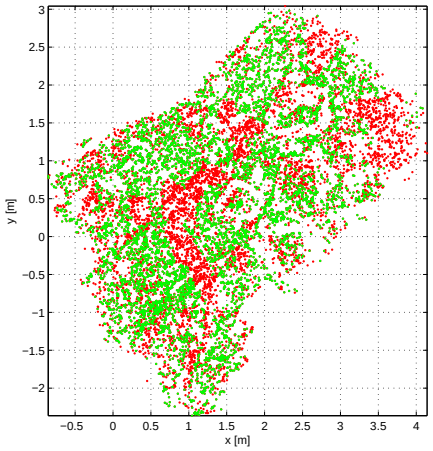


Fig. 20. Points below the median error (green) and above (red). Registration parameters where calculated using points below the median error. By referring to Figure 17 outliers tend to group around the smooth, raised folds of the carpet which clearly do not correspond to the drained carpet surface.

1999.

[8] H.S. Sawhney, S.C. Hsu, and R. Kumar. Robust video mosaicing through topology inference and local to global alignment. In *European Conference on Computer Vision*, pages 103–119, Freiburg, Germany, 1998.

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

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

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

[12] S.D. Fleischer, H.H. Wang, S.M. Rock, and M.J. Lee. Video mosaicking along arbitrary vehicle paths. In *Proceedings of the 1996 Symposium on Autonomous Underwater Vehicle Technology, 1996*, pages 293–299, Monterey, CA, June 1996.

[13] S. Negahdaripour, X. Xu, and L. Jin. Direct estimation of motion from sea fbor images for automatic station-keeping of submersible platforms. *IEEE Journal of Oceanic Engineering*, 24(3):370–382, July 1999.

[14] N. Gracias and J. Santos-Victor. Underwater mosaicing and trajectory reconstruction using global alignment. In *Oceans*, pages 2557–2563, Honolulu, Hawaii, 2001.

[15] O. Pizarro and H. Singh. Toward large-area underwater mosaicing for scientifi c applications. *IEEE Journal of Oceanic Engineering*, 28(4):651–672, 2003.

[16] R. Szeliski. Image mosaicing for tele-reality applications. Technical report crl 94/2, Cambridge Research Laboratory, Cambridge, MA, May 1994.

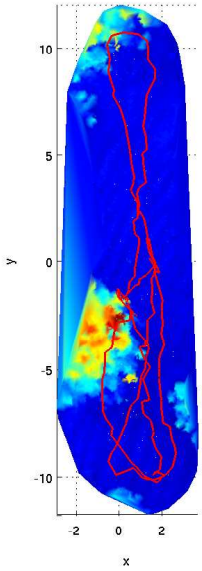
[17] C.C. Slama, editor. *Manual of Photogrammetry*. American Society of Photogrammetry, Bethesda, MD, fourth edition, 1980.

[18] S. Negahdaripour and H. , Madjidi. Stereovision imaging on submersible platforms for 3-d mapping of benthic habitats and sea-fbor structures. *IEEE Journal of Oceanic Engineering*, 28(4):625–650, October 2003.

[19] O. Pizarro, R. Eustice, and H. Singh. Relative pose estimation for instrumented, calibrated imaging platforms. In *Proceedings of the 2003 Conference on Digital Image Computing Techniques and Applications*, Sydney, Australia, 2003.

[20] R. Eustice, O. Pizarro, and H. Singh. Visually augmented navigation in an unstructured environment using a delayed state history. In *Accepted ICRA2004*, April 2004.

[21] S.C. Hsu and H.S. Sawhney. Influence of global constraints and lens distortion on pose and appearance recovery from a purely rotating



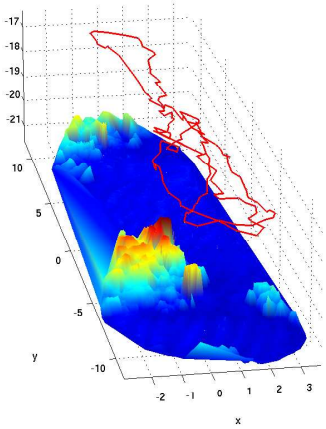


Fig. 21. Two views of the reconstruction as a surface through the recovered 3D points. The camera trajectory is also presented as a red line. Strong swell signifi cantly perturbed the vehicle trajectory yet the consistency of the reconstruction is apparent in the persistent features such as the sand ripples on the bottom.

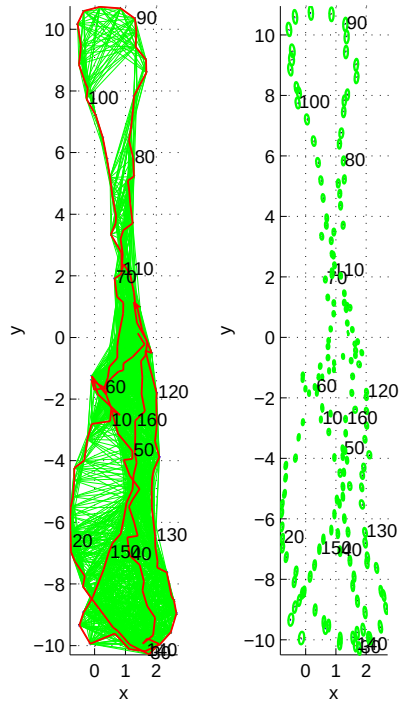


Fig. 22. (Left) Plan view of the camera trajectory (red) and common features between cameras (green links). (Right) The 99% confidence ellipses for the xy position of the cameras. Every tenth camera is numbered on both figures to suggest the temporal sequence

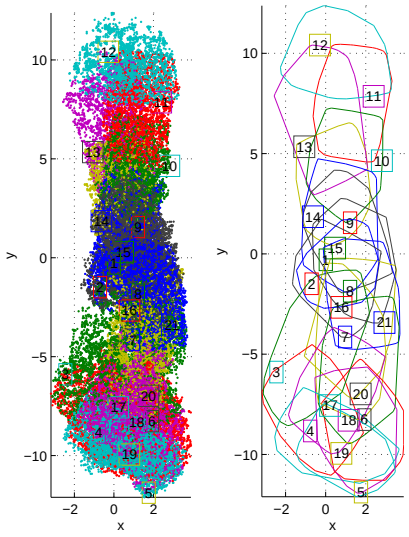


Fig. 23. (Left) Plan view of the features for each submap. (Right) Convex hull of the 3D features of each submap. The varying degrees of spatial overlap between submaps is apparent in these figures.

camera. In *Fourth IEEE Workshop on Applications of Computer Vision, 1998*, pages 154–159, October 1998.

[22] Z. Zhang and Y. Shan. Incremental motion estimation through local bundle adjustment. Technical Report MSR-TR-01-54, Microsoft Research, May 2001.

[23] P.A. Beardsley, A. Zisserman, and D. Murray. Sequential Updating of Projective and Affine Structure from Motion. *International Journal of Computer Vision*, 23(3):235–259, June 1997.

[24] C. Harris and M. Stephens. A Combined Corner and Edge Detector. In *Proceedings of the 4th Alvey Vision Conference*, pages 147–151, Manchester, U.K., 1988.

[25] T. Tuytelaars and L. Van Gool. Wide baseline stereo matching based on local, affinely invariant regions. In *Proceedings of the British Machine Vision Conference 2000*, pages 736–739, Bristol, UK, 2000.

[26] F. Mindru, T. Moons, and L. Van Gool. Recognizing color patterns irrespective of viewpoint and illumination. In *Proceedings of CVPR99*, pages 368–373, 1999.

[27] W.Y. Kim and Y.S. Kim. A region-based shape descriptor using Zernike moments. *Signal Processing:Image Communication*, 16(1-2):95–102, September 2000.

[28] O. Pizarro. *Large Scale Structure from Motion for Autonomous Underwater Vehicle Surveys*. Phd thesis, Massachusetts Institute of Technology and Woods Hole Oceanographic Institution, September 2004.

[29] M. A. Fischler and R Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[30] M. Bosse. *Atlas Framework for Scalable Mapping*. Phd, Massachusetts Institute of Technology, February 2004.

[31] P.J. Besl and N.D. McKay. A Method for Registration of 3-D Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-14(2):239–256, February 1992.

[32] B.K.P. Horn. Closed form solutions of absolute orientation using orthonormal matrices. *Journal of the Optical Society of America A*, 5(7):1127–1135, 1987.

[33] B. Triggs, P.F. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle adjustment - a modern synthesis. In Bill Triggs, A. Zisserman, and Robert Szeliski, editors, *Vision Algorithms: Theory & Practice*. Springer-Verlag, 2000.