2013

Hybrid Optical Acoustic Seafloor Mapping

Gabrielle Inglis
University of Rhode Island, gabrielle.inglis@gmail.com

HYBRID OPTICAL ACOUSTIC SEAFLOOR MAPPING

BY

GABRIELLE INGLIS

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

OCEAN ENGINEERING

UNIVERSITY OF RHODE ISLAND

2013

DOCTOR OF PHILOSOPHY DISSERTATION

OF

GABRIELLE INGLIS

APPROVED:

Dissertation Committee:

Major Professor
Chris Roman

Peter Swaszek

Harold Vincent

Nasser H. Zawia
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2013

ABSTRACT

Abstract

The oceanographic research and industrial communities have a persistent demand for detailed three dimensional sea floor maps which convey both shape and texture. Such data products are used for archeology, geology, ship inspection, biology, and habitat classification. There are a variety of sensing modalities and processing techniques available to produce these maps and each have their own potential benefits and related challenges. Multibeam sonar and stereo vision are such two sensors with complementary strengths making them ideally suited for data fusion. Data fusion approaches however, have seen only limited application to underwater mapping and there are no established methods for creating hybrid, 3D reconstructions from two underwater sensing modalities. This thesis develops a processing pipeline to synthesize hybrid maps from multi-modal survey data. It is helpful to think of this processing pipeline as having two distinct phases: Navigation Refinement and Map Construction. This thesis extends existing work in underwater navigation refinement by incorporating methods which increase measurement consistency between both multibeam and camera. The result is a self consistent 3D point cloud comprised of camera and multibeam measurements. In map construction phase, a subset of the multi-modal point cloud retaining the best characteristics of each sensor is selected to be part of the final map. To quantify the desired traits of of a map several characteristics of a useful map are distilled into specific criteria. The different ways that hybrid maps can address these criteria provides justification for producing them as an alternative to current methodologies. The processing pipeline implements multi-modal data fusion and outlier rejection with emphasis on different aspects of map fidelity. The resulting point cloud is evaluated in terms of how well it addresses the map criteria. The final hybrid maps retain the strengths of both sensors and show significant improvement over the single modality maps and naively assembled multi-modal

maps.

ACKNOWLEDGMENTS

The document that follows is the last piece of my grad school puzzle and I want to thank everyone who helped me put it together. We went some amazing places together, had some great laughs, broke things, fixed things, worked hard, played hard, learned more than we thought we could and generally had an incredible time. Without such great freinds and colleagues this research could not have taken shape, not to mention reached its final form.

I have to thank my advisor Chris Roman for many things but first and foremost, for taking a chance on me. He provided me with so many opportunities to learn and get involved in amazing projects in exciting parts of the world. His door was always open and he was willing to talk through any problem that had me stuck. I always walked away feeling inspired about my project and with at least six new ways of tackling whatever problem I was stuck on.

Ocean Engineering Department funded my first year giving me the chance to find my niche. For which I am so appreciative. I also owe much to the faculty of the OE and EE departments for the excellent instruction and guidence. Specifically Chris Baxter for helping me get here in the first place, as well as the members of my committee for their time and insight: Bud Vincent and Peter Swaszek. I am also appreciative of Gail's patience and knack for making sure things worked out.

The Ocean Exploration Trust provided the funding for much of my time at URI, but funding wasn't the only vital piece that they provided to my work. I want to thank all the talented people who have worked on the IFE/OET cruises over the years. Without them, there is no way I would have had such great data or amazing memories. Todd and Eric are amazing pilots, engineers and freinds. Sara has a story for any occasion I think we have spent most of our time together laughing. Kat helped me see the way through those first few cruises and is still a great friend and inspiration. Katy's good judgement and cool head kept things going smoothly and her amazing, upredictable sense of humor kept it light. I also don't think

I have ever beat her at backgammon. As the backbone of the operation Bob's enthusiasm for our mission has been contagious and has stuck with me through the years.

My amazing labmates ... there is no question that this thesis was possible in large part due to their inspiration, moral support and incredible skills. Without Ian's amazing software skills, and patient teaching, this project may well have taken years longer. It was great fun going to sea with Dave and I am pretty sure he can fix anything, well except one thing. Try jelly donuts next time, Dave. Clara was always there to help put things in perspective and talk about the bigger picture, or bikes. Bryan made sure I wasn't the only one walking into seminar 5 minutes late covered in mud and still wearing my mountain bike shoes, and still he managed to set the work-ethic bar higher than I ever thought possible.

There are so many other people that made my grad school life a memorable time. John was always up for every adventure, Jeannie's enthusiasm and sense of humor were balm for any crisis. I have never met more supportive women than Leigh, Kelly and Kelsey; I couldn't have asked for better people to have on my side down the home stretch.

Lastly I want to thank the friends and family far from Rhode Island who were supportive all through this project. Thanks especially to Allison and Natalie for sticking by me. I certainly owe my grandmother a nod: she did her PhD first, even when no other women did things like that. All my love and thanks goes my parents for taking midnight calls when I was sure I was stuck and everything was broken. They always knew I would get it working again and they always had my back. And Russ... Russ, thank you for trusting me and believing in me and being part of all the adventures of the last few years.

Its been a privilidge to have you all as part of my life and work.

Gabby

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

CHAPTER 1

Introduction

1.1 Problem Statement

The oceanographic research and industrial communities have a persistent need for high resolution three dimensional sea floor maps which convey both shape and texture. Specialized sea floor mapping techniques are used for marine archeology [1–4], marine geology [5–7], ship inspection [8, 9], and ecological monitoring [10–13]. Stereo cameras and multibeam sonars are among the instruments which can be used to make these maps. Both have their respective advantages and drawbacks. In the land robotics community, it is common to use complimentary mapping sensors and combine their measurements using data fusion techniques [14, 15]. This data fusion approach has seen only limited application to underwater mapping [16–18], and there are no established methods for creating hybrid, 3D reconstructions from two underwater sensing modalities. The goal of this thesis is to develop a method that integrates multibeam sonar and stereo vision data into a common navigation and mapping system to create hybrid maps. These hybrid maps serve the purpose of reducing multi-modality mapping data to an easily interpreted form while preserving as much detail as possible.

1.2 Background

Maps of marine environments provide vital insight for scientists and engineers. The last couple of decades have seen a boom in technologies which provide means to create increasingly accurate and detailed maps. These methods are based on a variety of sensing platforms including ships, tow-sleds, divers and increasingly underwater robots. A good overview of modern mapping platforms is given in [19].

Robotic platforms are of particular interest because they are able to make detailed observations of the underwater environment using a variety of sensors.

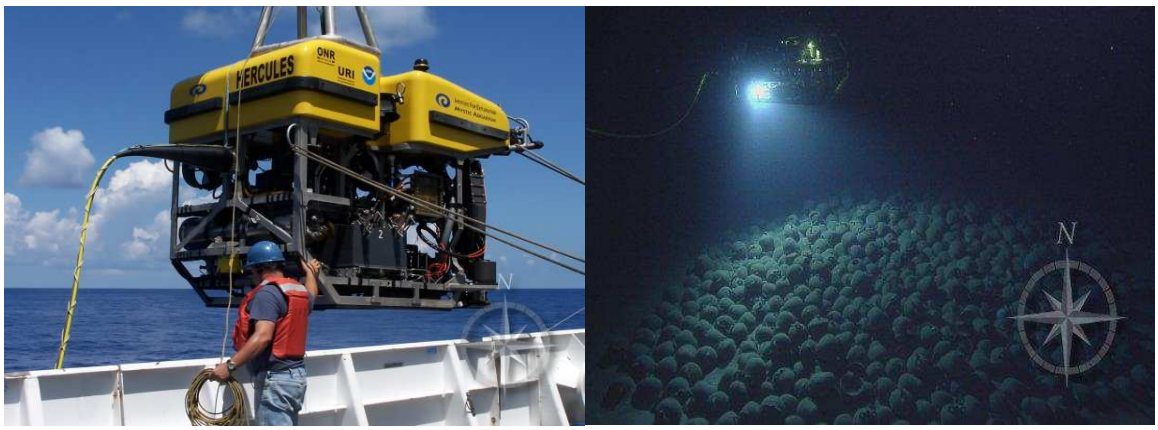


Figure 1. Remotely operated vehicle *Hercules*. Images of *Hercules* being deployed using a crane (left) and surveying a shipwreck (right).

They can travel to areas too deep, hot, or otherwise extreme for human divers and carry out more detailed and precise measurements than ship based platforms. They can acquire optical imagery of the sea floor which can't obtained from the surface due high rate of light attenuation in water.

These robots can be either Autonomous Underwater Vehicles (AUVs), which typically execute pre-programmed missions themselves, or ROVs which are controlled directly by a user throughout the course of the mission (Fig. 1). Mapping specific robots carry a suite of navigation sensors which measure depth, attitude, speed and relative position, as well as mapping sensors which make acoustic or optical range and backscatter measurements. They can also have sophisticated positioning and control systems in order to hold position or carry out structured surveys. Sites of interest can be traversed according to specific instructions to guarantee a certain amount of coverage (Fig. 2).

Robotic platforms are often equipped with several mapping modalities. This lends flexibility to the mapping system since each modality has its own strengths. The application of these sensors to navigation and mapping have been researched thoroughly by researchers in land robotics, computer vision, and photogrammetry. They are also well researched and understood in the underwater environment.

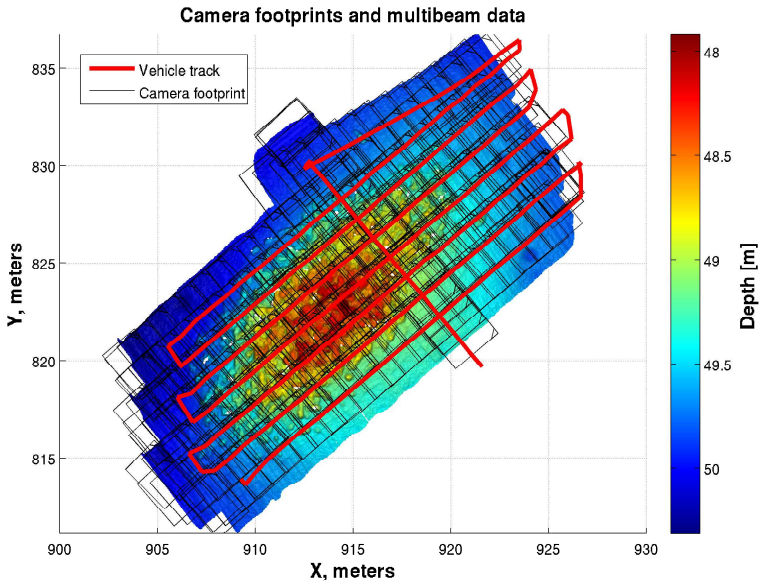


Figure 2. Survey tracklines. A typical underwater vehicle survey follows a back and forth "mowing the lawn" with trackline spacing that provides overlapping coverage for mapping sensors. The figure shows the tracklines in red, overlapping image footprints in red and a multibeam sonar map as the underlay. Notice the final trackline running orthogonal to the rest of the survey. This "loop closure", ensures that some terrain will be observed more than once in order to constrain drift in navigation.

Generally, a single survey will be executed using one instrument. Fusion of data from these sensors is just beginning to be explored in the underwater context. Kunz developed a system for navigation refinement using both sonar and camera [17]. Hurtos addressed the issue of finding the offsets between a camera system and a multibeam sonar which is critical to fusing their data [16]. To date however, there is no established method for created hybrid maps comprised of 3D structure from fused stereo and multibeam sonar range data.

This thesis develops a processing pipeline to synthesize hybrid maps from multi-modal survey data. It is helpful to think of this processing pipeline as having two distinct phases. The first is **navigation refinement**, where data from navigation sensors are corrected to enforce consistency between mapping measurements. The second phase is **map construction** where a selected subset of mapping data is projected into a common coordinate frame to construct a map.

1.3 Underwater Navigation

Sensor data is assembled into maps using robot position data. The most basic requirement of navigation data is that it is at least as good as the mapping sensors so that it does not become the dominant source of error in the map. Therefore, a lot of research has gone into improving underwater vehicle navigation so that it keeps pace with the improvements in mapping sensor resolution.

1.3.1 Dead Reckoning

Vehicle navigation data is acquired using a number of sensors to measure attitude, depth and velocity. The vector velocity can be is integrated over time to compute the distance traveled by the vehicle relative to its starting point. This relative position measurement can be combined with attitude and depth measurements can be used to form a 6 Degrees of Freedom (DOF) "dead reckoned" position estimate for the vehicle at any time. Because the random error in the measurements is also integrated over time, it accumulates and causes the trajectory estimate to drift from the true trajectory. Over the course of a survey, this type of error begins to dominate mapping sensor error. When the navigation data places the vehicle in an incorrect location, mapping measurements are also incorrectly localized. If the localization error is greater than the sensor precision, the measurements will appear misaligned.

1.3.2 Direct Measurement

Absolute position measurements such as GPS position are not available. So other methods have been developed to directly measure robot position and constrain navigation drift underwater. Long Baseline (LBL) acoustic systems are the underwater analog to GPS. A number of acoustic beacons are placed around a survey site. The vehicle is able to range to the beacons and determine its absolute position on the site. This approach has been used for underwater mapping in the detection of hydrothermal vent plumes as well as microbathymetric mapping of an

ancient shipwreck [20–22]. While high frequency LBL systems can localize with centimeter level accuracy without drift, the systems are expensive, cumbersome to deploy and not practical for all survey locations.

A less cumbersome alternative to LBL systems is Ultra Short Baseline (USBL) [23]. This method computes a position using range and bearing measurements from the ship to vehicle. While there is no drift, this method is susceptible to noise an typically offers accuracies between 0.1% and 1.5% of water depth.

1.3.3 Algorithmic Refinement

Algorithmic approaches provide an alternative to direct measurements for vehicle navigation refinement. The main advantages are that corrections can be applied in post processing and the methods do not rely on sensors external to the vehicle. In robotics, refining vehicle navigation is coupled with building a map of the environment. This is known as SLAM. SLAM utilizes the main tenants of estimation theory to maintain an estimate of the robot's pose history as well as a map of its surroundings with accuracy that exceeds dead reckoned vehicle position estimates. This is by done enforcing consistency between the robot trackline and locations of repeated observations. Recent advances in SLAM research make large scale and repeatable surveys more tractable than ever for all types of underwater surveys. SLAM is now considered a solved problem by its foremost researchers but there are still open research questions that arise in specific applications. For example, underwater navigation presents challenges such as survey size, unstructured scenes and limited features.

Several papers have addressed SLAM problems specific to underwater navigation and mapping. Roman creates a bathymetric map from multibeam sonar data by assembling sonar pings into submaps using a Kalman framework for filtering the navigation data. The submaps are used to constrain navigation drift when sections of the sea floor are re-observed [24, 25]. Eustice and Mahon both filter the navigation data using a pose based information filter and visual data to constrain

the robot pose [26, 27]. Barkby uses the principle of particle filtering to estimate robot pose and assemble a multibeam sonar based map [28, 29].

The SLAM examples mentioned above fuse data from multiple navigation sensors but only one mapping sensor to arrive at a final refined navigation solution which produces a self consistent map. In order to create a trajectory that maximizes consistency from two mapping sensors, both types of measurements must be used in the constraint network. Several examples of this have been investigated. Fallon fuses sidescan sonar and acoustic ranging within common navigation constraint network [30]. Hover et al and Kunz both solve for vehicle trajectories using constraints from multibeam sonar and monocular cameras [9, 17]. This concept is vital in aligning 3D structure from multibeam sonar and stereo cameras to create the proposed hybrid maps.

1.4 Mapping

The purpose of a map is to provide a meaningful reduction of the survey data for the end user. A scientist is interested in both structure and texture of the scene and makes interpretations based on colors, shapes, sizes, positions, so the map should be as metrically accurate as possible and easily texture mapped. In general, maps are composed of scene measurements projected into a common spatial frame of reference. There are two basic ways of producing maps for our purposes. The first is to optimize the map in SLAM or Structure from Motion (SFM) [31]. This means that navigation and mapping are simultaneously refined and the map is comprised of strictly the measurements used for navigation refinement. The second is to refine the vehicle poses using SLAM and then use the poses only to project environmental measurements into the common reference frame, this is known as *mapping with known poses* [32]. For the purposes of multi-modal mapping, the latter technique allows more flexibility to use instruments which don't lend themselves to probabilistic mapping easily and to create a map where two sensors can mutually reinforce each other [15].

1.4.1 Probabilistic Mapping

A straight forward approach to SLAM based mapping is to simply use SLAM solution's map. This map can be an occupancy grid where the map is a grid and cells are populated with the probability that something exists there [32] such as a hydrothermal vent plume [33]. Or it can be a sparse set of landmarks such as trees with compact descriptors [34]. These maps contain well localized information, but are highly abstracted. The level of abstraction is ideal for a autonomous mission planning but can be too abstract for detailed scientific inquiry.

A more informative map should be comprised of a 3D mesh or 2.5D gridded height map to convey the structure and a photo-realistic overlay to supply textural information. Barkby creates a probabilistic height map from multibeam sonar data alone using a non feature based particle filter [35]. SFM [31, 36, 37] approaches maintain a large enough number of sparse features that a detailed 3D mesh which can be easily texture mapped is a direct result. However, not all types of mapping data are suited to feature-based estimation frameworks. In particular sonar range data is more suited for pose based SLAM techniques. To ensure mutual alignment of the camera and the multibeam measurements, they must be incorporated into the same navigation refinement framework for estimating vehicle poses. This requires a navigation refinement system which is flexible enough to incorporate the feature based stereo constraints and pose based multibeam constraints. After the poses have been estimated, then it is possible to construct the map from the known poses.

1.4.2 Mapping with Known Poses

Mapping with known poses gives particular control over the map characteristics such as point density and blending techniques since the map is created independently of the navigation solution. When mapping with known poses, the map making process is distinct from the robot pose estimation process. There are a few instances of this technique being used in underwater mapping where

the primary goal is photo-realistic scene models. Johnson-Roberson creates photo-realistic scene models with high point density, using the poses found during a view-based SLAM solution [38, 39]. [40] estimates the camera trajectory and then computes a dense stereo correspondence map to create a dense scene reconstruction.

1.5 Justification for Use of Hybrid Maps

Currently there are no established methods which create hybridized 3D reconstructions from multibeam sonar and optical imagery for underwater mapping. However, since both modalities are readily available on mapping ROVs and AUVs, it is advantageous to present both data sets in a common mapping framework which incorporates the best attributes of each sensor. The complementary strengths of the sensors are related to operational range, scale, and spatial resolution. Their complimentary nature ideally suits them for fusion into a hybrid map to retain the best characteristics of each sensor.

Optimal operating scales vary between the two sensors. A multibeam sonar can be used to map at a wide range of scales, while a stereo camera system is far less flexible. The operational range of our particular multibeam is 1m to 20m, this coupled with its 90° swath width make it a useful tool for both micro-bathymetric $\mathcal{O}$(5cm) mapping and larger $\mathcal{O}$(km), less detailed surveys. On the other hand, the stereo camera system has an operational range of 1m to 4m with an across track field of view of 40°, so its utility is constrained to smaller areas.

While multibeam is useful for a variety of survey sizes and altitudes, stereo cameras have far greater spatial resolution. High spatial resolution, defined as the number of measurement points per unit area, is required to resolve small features. At 3m range, a 1.3 megapixel camera will have spatial resolution $\sim$ 40points/cm ^{2}. The multibeam sonar's 90° swath width is beamformed into 512 beams. At 3m of altitude, this translates into one measurement every centimeter. With a forward speed of 12cm per second and a ping rate of 12Hz, the along track resolution of

the multibeam is also 1 point per centimeter making the overall spatial resolution of the sensor $\sim 1\text{point}/\text{cm}^2$.

Camera and multibeam degrade under different circumstances. The stereo cameras are particularly sensitive to high turbidity, backscatter and limited texture. Suspended particles and sediments disturbed by currents or prop wash can obscure the site making it difficult or impossible to obtain usable range data from the cameras. In addition, the cameras are difficult to calibrate correctly. The calibration model used frequently in machine vision is violated underwater due to the differing indices of refraction between air and water. This modeling error is particularly apparent in highly structured scenes or at ranges that the calibration wasn't intended for [41]. This introduces warping to the final reconstruction. While turbidity and calibration are not a problem for the multibeam, the acoustic data can suffer from speckle noise, as well as artifacts introduced by the transducer geometry. These errors obscure fine details in the final map or cause holes in the mesh.

Using these two complimentary sensors together will lead to a more flexible survey apparatus, able to produce maps the highest possible quality available under any given set of conditions by leveraging each sensors respective strengths.

1.6 Contributions

The contributions of this project will be the following:

- A multi-modal navigation framework for simultaneously refining camera and multibeam poses throughout the survey using SLAM which emphasizes alignment between the two sensors.

- A mapping methodology which selects the best sensor data for each map location from a redundant data set while respecting the inherent characteristics of each sensor.

1.7 Assumptions

A number of assumptions are necessary to process the data used in this thesis.

- The approach utilizes navigation data that is adequate to constrain vehicle motion. Navigation data comes from a suite of on-board sensors which provide information regarding vehicle depth, attitude and velocity. This is enough to constrain the six degrees of freedom vehicle motion. Moreover, these or analogous sensors are present on nearly all underwater vehicles since they are required for basic functionality. Therefore, it is reasonable to assume that the navigation data for the surveys presented here and the vast majority of underwater surveys will be adequate to constrain vehicle motion.

- A constant velocity model is adequate to predict vehicle motion between navigation sensor measurements. The constant velocity assumption has been used previously with good results for filtering underwater vehicle motion [42]. This is because sea floor survey type missions are executed slowly and without abrupt changes to vehicle speed or attitude which might disrupt the mapping data quality. Additionally, vehicles capable of imaging surveys are generally passively stable in pitch and roll, making the platform unlikely to violate the constant velocity assumption by abrupt attitude changes. Measurements are also made very frequently so relative motion between measurements is small.

- Estimates of stereo camera calibration and sensor offsets are available. Camera calibrations must be done in order to obtain any metric information from a camera and can be obtained using a variety of methods before the sensors are taken into the field. The sensor offsets are straightforward to obtain by hand measuring relative to the vehicle pose.

- An ideal pinhole camera model is valid over a given survey. While standard cameras don't precisely conform to a pinhole model underwater, radial distortion parameters computed during camera calibration closely approximate the

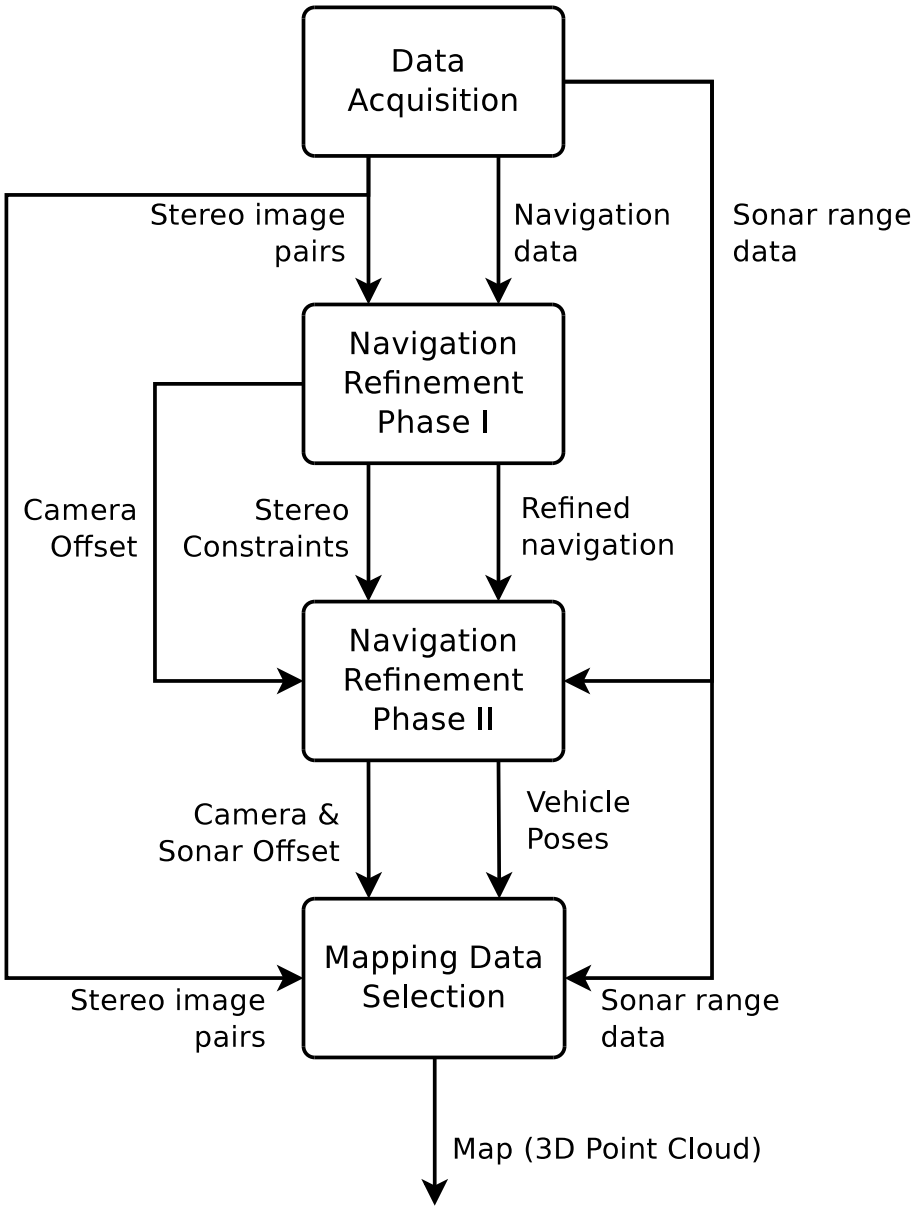
effect. This assumption holds approximately if the camera is positioned close to the viewport glass of its housing and the camera altitude stays relatively close to its calibrated altitude [41]. The cameras used for this survey were mounted near the viewport glass with the former constraint in mind. The latter constraint is addressed by the nature of underwater optical surveys. Surveys typically have a constant moderate altitude since high altitudes preclude imaging due to rapid light attenuation and low altitudes make only a small amount of terrain observable at once. A small amount of error may result from this assumption, but the advantages gained in efficiency by leveraging pinhole camera geometry and constraints are substantial enough to outweigh it.

- Overlap exists for some of the mapping sensor measurements to provide constraints on vehicle pose and sensor offset estimation. Surveys can generally be designed to incorporate as much overlap as desired. The survey design trades off available time with size of area covered and instrument field of view. Sometimes a survey mission will be aborted before a final loop can be closed, but this can be mitigated by building overlap into the body of the survey and not just relying on a single overlapping trackline at the end.

1.8 Layout

The process of map making in this thesis has two parts. The first part is navigation refinement where navigation and mapping sensor data are combined to estimate refined vehicle poses. The second part uses the poses determined in part one to assemble a map. The map construction uses the inherent characteristics of each sensor to reject outliers and choose the best sensor for each section of the map. A detailed chart showing how the various processing steps are related is in (Fig. 1.8).

Chapter 2 describes navigation refinement. In this section, the vehicle posi-


Data Acquisition
Stereo image pairs
Navigation data
Sonar range data
Navigation Refinement Phase I
Camera Offset
Stereo Constraints
Refined navigation
Navigation Refinement Phase II
Camera & Sonar Offset
Vehicle Poses
Mapping Data Selection
Stereo image pairs
Sonar range data
Map (3D Point Cloud)

tion at each mapping sensor measurement is estimated. Constraints on the vehicle trajectory are derived from multiple observations of the same terrain by the mapping sensors. The trajectory estimation solves for the vehicle poses which best explain the constraints. Chapter 3 describes map construction. The map is assembled by projecting measurements into a common reference frame and then culling redundant data and outliers. The final Chapter contains so resulting maps and discussion.

List of References

[1] O. Pizarro and H. Singh, "Toward large-area mosaicing for underwater scientific applications," *IEEE Journal of Oceanic Engineering*, vol. 28, no. 4, pp. 651–672, October 2003.

[2] R. Ballard, L. Stager, D. Master, D. Yoerger, D. Mindell, L. Whitcomb, H. Singh, and D. Piechota, "Iron age shipwrecks in deep water off Ashkelon, Israel," *American Journal of Archeology*, vol. 106, no. 2, April 2002.

[3] I. Mahon, O. Pizarro, M. Johnson-Roberson, A. Friedman, S. Williams, and J. Henderson, "Reconstructing pavlopetri: Mapping the world's oldest submerged town using stereo-vision," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, 2011, pp. 2315–2321.

[4] H. Singh, J. Adams, D. Mindell, and B. Foley, "Imaging underwater for archaeology," *Journal of Field Archaeology*, vol. 27, no. 3, pp. 319–328, 2000.

[5] H. Sigurdsson, S. Carey, M. Alexandri, G. Vougioukalakis, K. Croff, C. Roman, D. Sakellariou, C. Anagnostou, G. Rousakis, C. Ioakim, A. Goguo, D. Ballas, T. Misaridis, and P. Nomikou, "Marine investigations of greece's santorini volcanic field," *Eos, Transactions American Geophysical Union*, vol. 87, no. 34, pp. 337–342, 2006. [Online]. Available: http://dx.doi.org/10.1029/2006EO340001

[6] D. Yoerger, A. Bradley, B. Walden, M.-H. Cormier, and W. Ryan, "Fine-scale seafloor survey in rugged deep-ocean terrain with an autonomous robot," in *Robotics and Automation, IEEE International Conference on*, vol. 2, 2000, pp. 1787–1792.

[7] V. Brandou, A. G. Allais, M. Perrier, E. Malis, P. Rives, J. Sarrazin, and P. M. Sarradin, "3D reconstruction of natural underwater scenes using the stereovision system IRIS," in *IEEE OCEANS '07*, Aberdeen, 2007, pp. 1–6.

[8] S. Negahdaripour and P. Firoozfam, "An ROV stereovision system for ship hull inspection," pp. 551–564, 2006.

[9] F. Hover, R. Eustice, A. Kim, B. Englot, H. Johannsson, M. Kaess, and J. Leonard, "Advanced perception, navigation and planning for autonomous in-water ship hull

inspection," *Intl. J. of Robotics Research, IJRR*, vol. 31, no. 12, pp. 1445–1464, Oct 2012.

[10] V. E. Kostylev, B. J. Todd, G. B. J. Fader, R. C. Courtney, G. D. M. Cameron, and R. A. Pickrill, "Benthic habitat mapping on the Scotian Shelf based on multibeam bathymetry, surficial geology and sea floor photographs," *Marine Ecology Progress Series*, vol. 219, pp. 121–137, Sep 2001.

[11] M. Johnson-Roberson, S. Kumar, O. Pizarro, and S. Willams, "Stereoscopic imaging for coral segmentation and classification," in *IEEE OCEANS '06*, Sept 2006, pp. 1–6.

[12] S. Williams and I. Mahon, "Simultaneous localisation and mapping on the Great Barrier Reef," in *Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on*, vol. 2, May 2004, pp. 1771 – 1776 Vol.2.

[13] A. Kenny, I. Cato, M. Desprez, G. Fader, R. Schüttenhelm, and J. Side, "An overview of seabed-mapping technologies in the context of marine habitat classification," *ICES Journal of Marine Science: Journal du Conseil*, vol. 60, no. 2, pp. 411–418, 2003.

[14] B. Douillard, D. Fox, F. Ramos, and H. Durrant-Whyte, "Classification and semantic mapping of urban environments," *The International Journal of Robotics Research*, vol. 30, no. 1, pp. 5–32, January 2011.

[15] P. Gurram, H. Rhody, J. Kerekes, S. Lach, and E. Saber, "3d scene reconstruction through a fusion of passive video and lidar imagery," in *Applied Imagery Pattern Recognition Workshop, 2007. AIPR 2007. 36th IEEE.* IEEE, 2007, pp. 133–138.

[16] M. Hurts, X. Cuf i Soler, and J. Salvi, "Integration of optical and acoustic sensors for 3d underwater scene reconstruction." *Instrumentation ViewPoint*, no. 8, pp. 43–, 2009. [Online]. Available: http://dialnet.unirioja.es/servlet/articulo?codigo=3201922

[17] C. Kunz, "Autonomous underwater vehicle navigation and mapping in dynamic, unstructured environments," Ph.D. dissertation, MIT-WHOI Joint Program, November 2011.

[18] S. Negahdaripour, H. Sekkati, and H. Pirsiavash, "Opti-acoustic stereo imaging: on system calibration and 3-d target reconstruction," *Trans. Img. Proc.*, vol. 18, no. 6, pp. 1203–1214, June 2009. [Online]. Available: http://dx.doi.org/10.1109/TIP.2009.2013081

[19] L. A. Mayer, "Frontiers in seafloor mapping and visualization," *Marine Geophysical Researches*, vol. 27, no. 1, pp. 7–17, 2006.

[20] M. V. Jakuba, C. N. Roman, H. Singh, C. Murphy, C. Kunz, C. Willis, T. Sato, and R. A. Sohn, "Long-baseline acoustic navigation for under-ice autonomous underwater vehicle operations," *Journal of Field Robotics*, vol. 25, no. 11-12, pp. 861–879, 2008. [Online]. Available: http://dx.doi.org/10.1002/rob.20250

[21] D. R. Yoerger, M. Jakuba, A. M. Bradley, and B. Bingham, "Techniques for deep sea near bottom survey using an autonomous underwater vehicle," *The International Journal of Robotics Research*, vol. 26, no. 1, pp. 41–54, 2007.

[22] 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.

[23] P. Rigby, O. Pizarro, and S. Williams, "Towards geo-referenced auv navigation through fusion of usbl and dvl measurements," in *OCEANS 2006*, 2006, pp. 1–6.

[24] C. N. Roman, "Self consistent bathymetric mapping from robotic vehicles in the deep ocean," Ph.D. dissertation, MIT/WHOI Joint Program, 2005.

[25] C. Roman and H. Singh, "A Self-Consistent bathymetric mapping algorithm," *Journal of Field Robotics*, vol. 24, no. 1-2, pp. 23–50, 2007.

[26] S. Williams, O. Pizarro, I. Mahon, and M. Johnson-Roberson, "Simultaneous localisation and mapping and dense stereoscopic seafloor reconstruction using an auv," in *Experimental Robotics*, ser. Springer Tracts in Advanced Robotics, O. Khatib, V. Kumar, and G. Pappas, Eds. Springer Berlin / Heidelberg, 2009, vol. 54, pp. 407–416.

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

[28] S. Barkby, S. Williams, O. Pizarro, and M. Jakuba, "An efficient approach to bathymetric slam," in *2009 IEEE/RSJ International Conference on Intelligent robots and systems, Proceedings on*. Piscataway, NJ, USA: IEEE Press, 2009, pp. 219–224.

[29] S. Barkby, S. Williams, O. Pizarro, and M. Jakuba, "Incorporating prior maps with bathymetric distributed particle slam for improved auv navigation and mapping," in *Proceedings of OCEANS 2009, MTS/IEEE Biloxi*. IEEE Press, Oct. 2009, pp. 1–7.

[30] M. F. Fallon, M. Kaess, H. Johannsson, and J. J. Leonard, "Efficient AUV navigation fusing acoustic ranging and side-scan sonar," in *2011 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, May 2011, pp. 2398–2405.

[31] O. Pizarro, R. Eustice, and H. Singh, "Large area 3-d reconstructions from underwater optical surveys," *Oceanic Engineering, IEEE Journal of*, vol. 34, no. 2, pp. 150–169, 2009.

[32] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. The MIT Press, September 2005.

[33] M. Jakuba and D. Yoerger, "Autonomous search for hydrothermal vent fields with occupancy grid maps," in *Proc. of ACRA*, vol. 8, 2008, p. 2008.

[34] F. Dellaert, "Factor graphs and gtsam: A hands-on introduction," 2012.

[35] S. Barkby, S. B. Williams, O. Pizarro, and M. V. Jakuba, "A featureless approach to efficient bathymetric slam using distributed particle mapping," *Journal of Field Robotics*, vol. 28, no. 1, pp. 19–39, 2011.

[36] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski, "Building rome in a day," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 72–79.

[37] T. Nicosevici, N. Gracias, S. Negahdaripour, and R. Garcia, "Efficient three-dimensional scene modeling and mosaicing," *Journal of Field Robotics*, vol. 26, no. 10, pp. 759–788, 2009. [Online]. Available: http://dx.doi.org/10.1002/rob.20305

[38] M. Johnson-Roberson, O. Pizarro, S. Williams, and I. Mahon, "Generation and visualization of large scale 3D reconstructions from underwater robotic surveys," *Journal of Field Robotics*, 2009 (in press).

[39] I. Mahon, S. Williams, O. Pizarro, and M. Johnson-Roberson, "Efficient view-based SLAM using visual loop closures," *Robotics, IEEE Transactions on*, vol. 24, no. 5, pp. 1002–1014, Oct. 2008.

[40] A. Sedlazeck, K. Koser, and R. Koch, "3d reconstruction based on underwater video from rov kiel 6000 considering underwater imaging conditions," in *OCEANS 2009-EUROPE*. IEEE, 2009, pp. 1–10.

[41] T. Treibitz, Y. Schechner, C. Kunz, and H. Singh, "Flat refractive geometry," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, May 2011, PMID: 21576744. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/21576744

[42] C. Roman and H. Singh, "Improved vehicle based multibeam bathymetry using sub-maps and slam," in *IROS'05: Proceedings of the 2005 IEEE/RSJ international conference on Intelligent robots and systems*, 2005, pp. 3662–3669.

CHAPTER 2

Multibeam and Stereo SLAM

2.1 Introduction

Navigation refinement is a crucial step in underwater mapping. Dead reckoned vehicle positions alone accumulate error that grows unbounded with time and causes misalignment between mapping measurements. The goal of the navigation refinement step is to reduce navigation error until it is no longer the dominant source of error in the map. Several approaches, both instrument based and algorithmic were discussed in the Chapter 1. This Chapter begins by explaining several algorithmic approaches that have been used to incorporate data from multibeam and cameras into a refined vehicle trajectory. Second, the existing work is extended to incorporate methods which increase measurement consistency between both multibeam and camera. Then a set of error metrics are summarized which help evaluate the utility of the method. Finally, results are presented and evaluated using the proposed error metrics.

2.2 Background

2.2.1 Filtering SLAM: Submap SLAM and SEIF SLAM

The Extended Kalman Filter (EKF) has been a common tool for navigation refinement since Smith Self and Cheeseman advocated its use for building probabilistic maps in the 1980's [1]. This filtering approach was applied to underwater mapping by Roman to assemble multibeam bathymetric maps [2]. The key aspect to Roman's implementation is assembling adjacent sonar pings into submaps in which navigation drift contributes less error than sensor resolution, and can therefore be neglected. Navigation data and uncertainties are accumulated in the EKF which augments the filtered vehicle state vector with delayed states corresponding to locations of the submap origins. Links between overlapping submaps are made when the structure of the two submaps can be registered. Relative poses between

submap origins added to the filter as additional measurements between the delayed states to produce a well constrained vehicle trajectory that corresponds to a self consistent map. The utility of this method is limited however by its $\mathcal{O}(n^3)$ complexity where n is the size of state space. As a result it is impractical for refining trajectories with many unknown submap origins or image poses. [2].

The Sparse Extended Information Filter (SEIF) has been used as an alternative to the EKF because it scales well in state space. The information matrix of the filter is maintained instead of the covariance matrix so that the update step does not require an $\mathcal{O}(n^3)$ inversion. Additionally, the information matrix is exactly sparse when the variables to be estimated consist of prior poses alone, a characteristic which can be exploited for efficient state recovery with $\mathcal{O}(n)$ complexity [3]. This method has been applied to underwater mapping with both monocular and stereo vision by Eustice [4] and Mahon [5] respectively.

Filtering leaves a few issues unresolved. EKFs and SEIFs estimate the current robot pose by applying a recursive filter to the previous pose, current measurements and control inputs. Because of this sensor updates only propagate forward so at updates with large measurement innovations the trajectory can become less smooth than might naturally be expected. Furthermore, linearization error can accumulate over the course of a trajectory.

2.2.2 Smoothing versus filtering

Recently, another approach known as Smoothing and Mapping (SAM) has been applied to address the lingering problems associated with filtering. In smoothing, the robot trajectory is not marginalized out and inference is done on the entire trajectory [6]. Since the full non-linear problem can be solved over the entire trajectory, error is evenly distributed around the graph. This produces a trajectory that is consistent with all of the constraints. It also produces smoother trajectories than filtering methods leading to more appealing maps [7].

Smoothing treats the SLAM problem as a large non-linear system which is

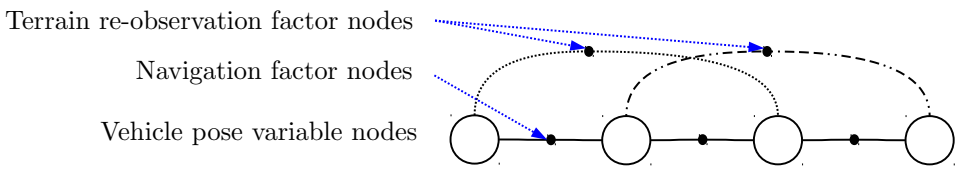


Figure 3. A SLAM Factor Graph. This factor graph contains variable nodes which are unknown vehicle poses and factor nodes which are measurements that act as constraints on the variable nodes. Constraints on temporally adjacent variable nodes are based on the vehicle navigation sensors and terrain re-observation factors are based on multiple measurements of the same terrain from mapping sensors.

solved all at once. It contains unknowns such as vehicle poses or landmark positions which are a function of measurements such as range and bearing to a landmark or vehicle velocity. This problem can be posed as as a factor graph [6]. This graphical model is an intuitive way to look at the system, breaking it down into variable nodes (variables to be estimated) and factor nodes (measurement functions)(Fig. 3). The goal is to find the Maxiumum a Posteriori (MAP) estimate for the unknowns. Ultimately the graph or non-linear system can be solved using a variety of inference methods. For most practical situations, the sparsity of the underlying structure allows for a solution using sparse matrix techniques which are highly efficient.

Smoothing algorithms are traditionally non-causal which generally precludes real time applications. However, the development of incremental Smoothing and Mapping (iSAM) gives an efficient method for incrementally adding new measurements in real time while keeping the vehicle position estimate current. As a result, smoothing approaches are currently being applied to underwater water robotics for both real time and post processed navigation and mapping. Hover et al combine an imaging sonar and monocular camera constraints incrementally within a factor graph to navigate and build a map during ship hull inspection [8]. Kunz uses multibeam sonar and stereo cameras on an AUV to build a map of a coral reef and refine sensor offsets for biological monitoring [9]. The results in both cases are robot trajectories which obey constraints imposed by mapping sensors and navigation sensors and maps which appear self-consistent..

A recent underwater mapping method combines both filtering and smoothing techniques. The EKF is useful for creating submaps for data association and map assembly, but scaling limitations make it impractical for refining long trajectories [10]. To mitigate the scaling issue, Vaughn leverages EKF assembled submaps for data association, but uses submap origins as nodes in a factor graph. The factor graph is then solved using the iSAM software package. This avoids scaling issues and efficiently improves navigation for map making [7].

2.3 Methods

The following navigation refinement methods are designed to estimate vehicle poses using data from both the on-board navigation and mapping sensors. They extend the state of the art in navigation refinement to enforce consistency between multiple modalities. The estimated poses will be used in the mapping phase (Chapter 3) to project measurements into a common reference frame and assemble a map from two sensors. Camera data is incorporated by aligning overlapping sets of images. For multibeam, it is common to aggregate sets of pings into submaps which can be aligned using point cloud registration techniques.

The process detailed in the following section is founded on the work of Kunz [9]. That procedure begins by aligning overlapping sets of images and incorporating them into a bundle adjustment style navigation solution where consistency is enforced using multiple views of the same landmark [11]. The resulting navigation solution is used to assemble multibeam submaps and establish links between those that overlap. A final navigation solution is then estimated using all of the available constraints, both camera and multibeam. This thesis adds a new step where overlapping camera and multibeam submaps are co-registered to refine their relative pose and enforce mutual consistency. Additionally, this approach is able to estimate the offsets of the sensors with respect to the vehicle frame as part of the navigation solution.

The process is outlined in Figure 4. This figure explicitly breaks the process

into two phases. The first phase refines camera offset and vehicle navigation. From this phase, multibeam constraints can be computed and further refinement of navigation data along with multibeam sensor offset. The specifics elements of this chart will be further explained in the following sections.

2.3.1 Instrumentation and platform

The data for this work was gathered during surveys using the ROV *Hercules* (Fig. 5). The data sets were collected during the 2012 field season in the Aegean Sea. Dense gridded surveys on spatial scales $\mathcal{O}(100\text{m}^2)$ were designed and executed over these archeological sites in order to gather simultaneous acoustic and optical imagery with approximately 200% overlap both along-track and across-track. An outline of the navigation and survey instruments available on the *Hercules* ROV is presented in Tables 1 and 2.

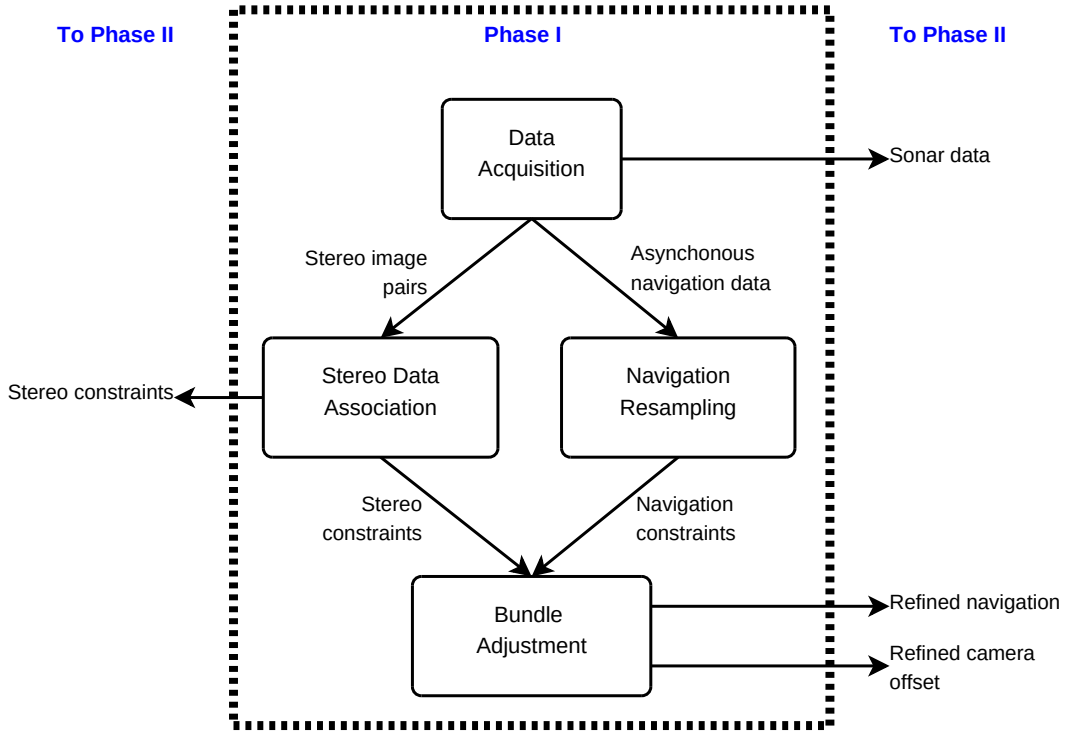
Table 1. Navigation sensors

Measurement	**Sensor**	**Precision**
Heading (north seeking)	OCTANS FOG	$\pm.1^\circ$
Pitch/Roll	OCTANS	$\pm0.01^\circ$
Depth (surface relative)	Pressure sensor	±0.01m
Velocity (bottom relative)	Acoustic Doppler (DVL)	±0.01m/s

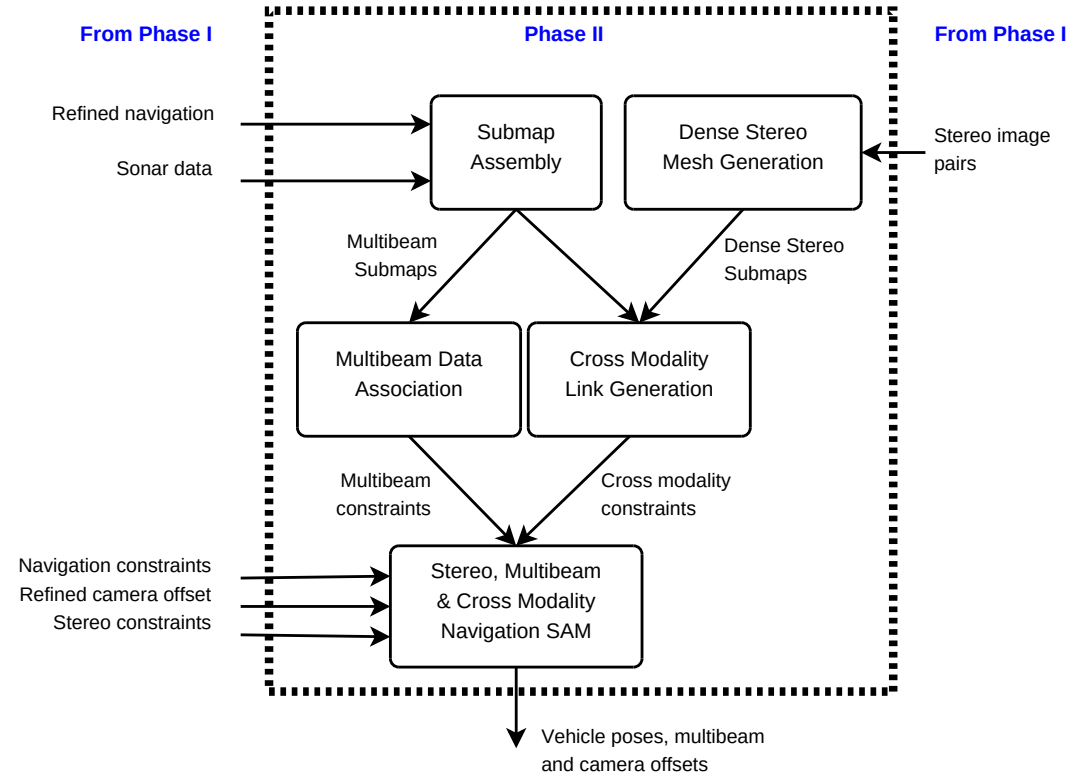
Table 2. Mapping sensors

Measurement	**Sensor**	**Precision**
Optics (Cameras)	Prosilica GC1380 BW Prosilica GC1380C Color	12-bit images 1360×1024 format
Acoustics (Multibeam)	Blueview MB1350	$\sim 1\%$ of range 512 beams

Acoustic data was collected using a Blueview MB1350 multibeam sonar. This is a particularly high frequency 1.35 MHz system with a 90° field of view. For typical surveys this translates to approximately 4-6 m swath widths. The instrument is mounted at the lowest aft point bringing it as close as possible to the sea floor



(a) Phase I



(b) Phase II

Figure 4. Flowchart of navigation refinement steps.

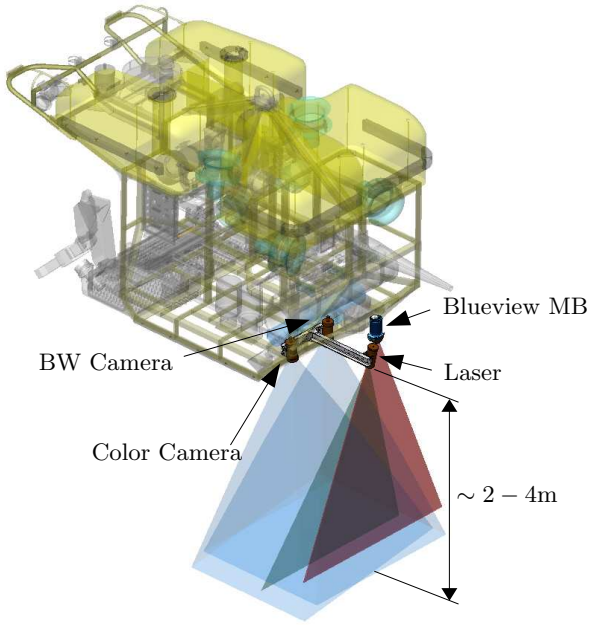


Figure 5. The *Hercules* ROV with stereo cameras and Blueview multibeam sonar. Flashbulb strobes are located on the forward section of the vehicle.

while the ROV maintains a minimum safe survey altitude of 2m. This allowes us to take advantage of the greater resolution available at reduced range.

Images were taken using a rigid stereo rig fitted with two Prosilica GC1380 cameras. These were mounted within pressure housings with flat glass viewports 300mm apart. Their optical axes are parallel. The color and black and white images were acquired as 12 bit grayscale and 48 bit Bayer respectively by 1024 by 1360 pixel CCDs.

The lighting was supplied using two Ocean Imaging Systems model M3831 flashbulbs hardware triggered off the master camera. These were mounted on the forward half of the vehicle to minimize the common volume of water imaged by the cameras and strobes. The maximum framerate that could be achieved was limited by the strobe recharge time to about 0.125Hz which translates to 1.25 frames per meter of travel along track.

Hercules's navigation instrumentation includes an Ixsea Octans fiber optic

gyro, a Paro Scientific depth sensor, and a Teledyne RDI doppler velocity log. The navigation data has several applications. The measurements are processed in real-time to drive the vehicle's autopilot allowing precise survey patterns. It is also visualized and logged using DVLNav software [12] for use in post processing.

2.3.2 Notation

It is helpful to specify a notation system for coordinate system transforms which indicate how constraints are integrated into the factor graph. This notation helps to articulate the spatial relationships formed by the network of constraints.

Coordinate systems

There are several relevant coordinate reference frames which will be referred to frequently (Fig. 6). The local level coordinate system, ℓ is an absolute frame. Its origin is in one fixed location. For convenience the origin of ℓ is assigned as the pose of the vehicle at t_0. The vehicle coordinate system v has its origin at a fixed location on the front of the vehicle. The term vehicle pose refers to the position and orientation of v within ℓ. The sensor coordinate frames are specific to each sensor. Multibeam sonar frame m and the left camera c are the sensors referred to most frequently. The right camera is offset from the left camera using a transform determined during stereo calibration. By convention however, the camera based 3D point clouds are expressed in left camera coordinate system. The position of these sensors within v is defined by the sensor offsets $(\mathbf{o}_{c,v},\ \mathbf{o}_{m,v})$,. The sensor offsets are rigid transformations which can be measured by hand on the vehicle and will be refined during navigation refinement.

The position and attitude of the vehicle with respect to ℓ at time i is $\mathbf{x}_{i,\ell}$. The odometry between vehicle poses at time i and j is written as $\mathbf{x}_{i,j}$ which is a transform that can be computed using the operations described in the following section.

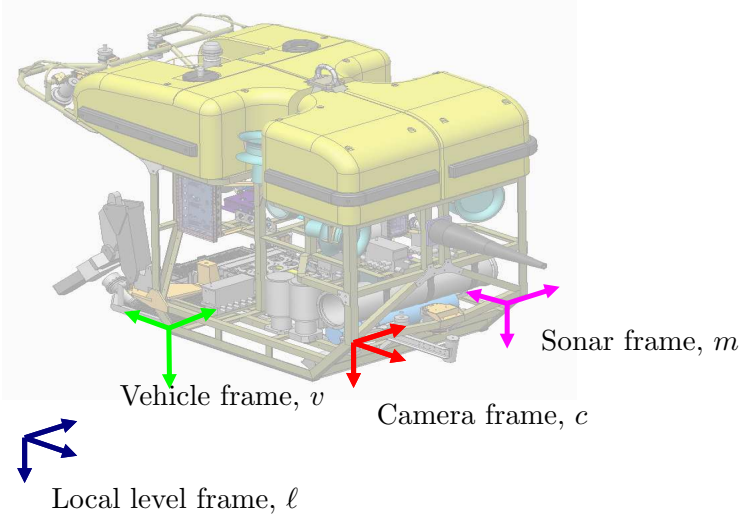


Figure 6. Relevant vehicle coordinate reference frames. The three main types of coordinate frames are the world frame w, the vehicle frame v and the sensor frames, m and c. Note that all frames have the z axis pointing down and vary based on the orientation of x and y

Spatial relationships between coordinate systems

A robot or sensor's position and attitude with respect to any coordinate system can be described in terms of the spatial variables $x, y, z, \theta, \phi, \psi$. The first three are translational variables and the last three are the attitude variables indicating roll, pitch and yaw respectively.

Operations on these variables allow a given pose to be expressed in other coordinate frame. The notation adopted for coordinate transforms is fully explained in Smith Self and Cheeseman [1]. However, the relevant transforms are summarized here. The compounding operation takes two relationships $\mathbf{x}_{i,j}$ and $\mathbf{x}_{j,k}$ and lays them head to tail to arrive at the compound relationship $\mathbf{x}_{i,k}$ It is known as the head-to-tail operation and is expressed as $\oplus$. The inverse relationship is useful as well. This might be used to reverse a spatial relationship that has been applied and is expressed as $\ominus$. A composite relationship known as tail-to-tail is useful for finding the relative pose between two forward relationships. The tail-to-tail is expressed as $\mathbf{x}_{j,k} = \ominus\mathbf{x}_{i,j} \oplus \mathbf{x}_{i,k}$. These operations offer a way to express the changes in spatial relationships between coordinate systems which occur due to

vehicle motion and sensor measurements.

2.3.3 Factor graph assembly and structure

A factor graph is a graphical model which expresses a large function in terms of its factors. It is intuitive to look at and can be solved using a variety of Bayesian inference methods. The goal of the navigation refinement using factor graphs is to determine the position of the mapping sensors at the time of measurement. A factor graph is used to structure the network of constraints from which poses will be inferred. One group of constraints consists of the dead reckoned navigation between these poses. Images can be abstracted into features and aligned to provide additional visual constraints. Multibeam pings can be assembled into submaps and aligned with each other to provide further constraints. Finally, alignments between stereo pair reconstruction and multibeam submaps enforce alignment between the two modalities using a third type of constraint..

The factor graph is assembled and solved in two phases (Fig. 4). In Phase I the feature based links and navigation links are used to solve for the vehicle positions and the offset of the cameras. Then the offset of the camera is held fixed and a graph containing the feature based constraints between cameras, multibeam constraints, and cross modality constraints is solved to find the vehicle poses and multibeam offset.

Computing navigation constraints between sequential mapping sensor measurements

The navigation data from the depth, attitude and velocity sensors provides constraints between sequential mapping sensor measurements (Fig. 7). However, the navigation data is asynchronous with the mapping sensor measurements, and must be resampled. The resampling is done using an EKF. Each successive measurement from a navigation sensor is incorporated into the filter using an update step. A prediction step is run when a multibeam ping or image capture step occurs

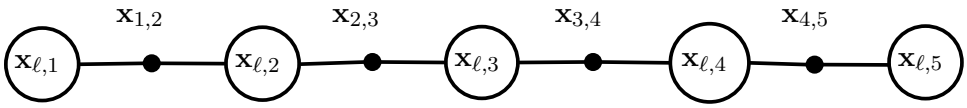


Figure 7. Factor graph with navigation based factors. These factor nodes only constrain temporally adjacent nodes and do not prevent drift in navigation data.

in order to recover the vehicle state and state covariance at that time. The relative poses and covariances between sequential mapping measurements are retained as constraints for the factor graph. These constraints only link temporally adjacent vehicle poses and will contain dead reckoning error which accumulates over time, necessitating the other forms of constraint (Fig. 7).

Data association

Data association is the process of recognizing that two separate observations relate to the same terrain and deriving a spatial constraint from their relative alignment. Depending on the sensor, there are two possible approaches to data association.

The first approach is for creating links between stereo image pairs. Linking stereo camera poses requires abstracting images into matchable features and recognizing a link between two poses when a unique feature is viewed in both poses. The second is for creating links between 3D terrain patches, generated using either camera or multibeam. Establishing links between 3D patches is done by aligning the structure of two overlapping patches using point cloud registration techniques. This approach is appropriate for both multibeam-multibeam links and multibeam-stereo cross modality links (Fig. 8).

Generally when SLAM algorithms are performed online, links are sequentially hypothesized when a measurement from an adjacent pose lies within the a confidence ellipse related to the covariance of the current pose or current measurement. The covariance is kept small because the navigation is continually being refined. This results in a small search area for potential links and a robustness to false

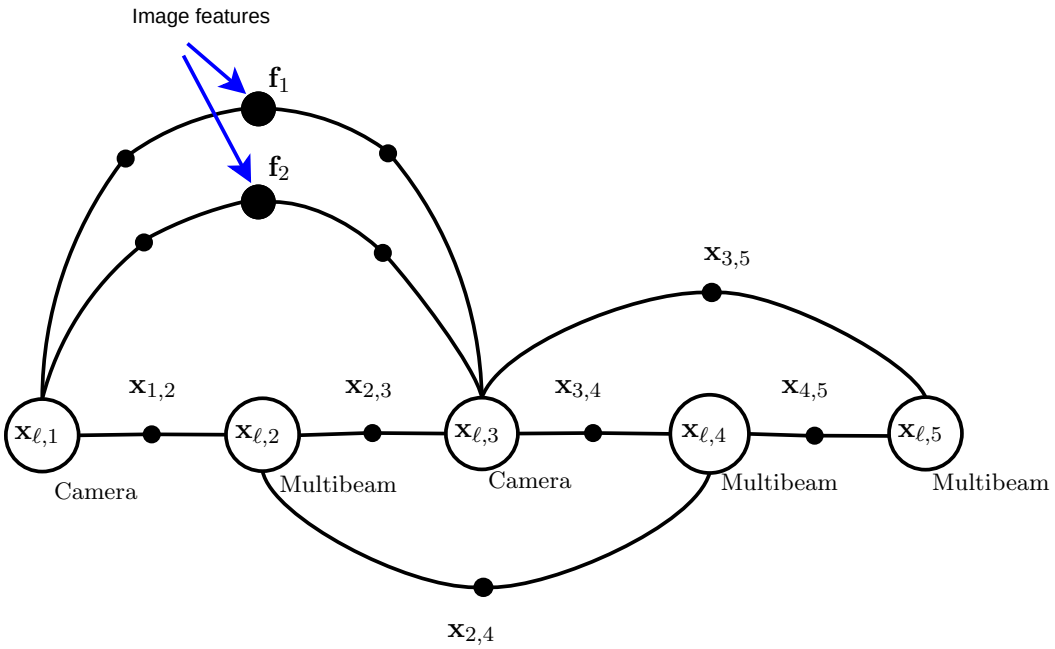


Figure 8. Factor graph with navigation and data association factor nodes. The data association further constrains the navigation by creating constraints between spatially adjacent pose nodes. The image-based links between poses 1 and 3 are based on re-observation of the same two image features $\mathbf{f}_1$ and $\mathbf{f}_2$. This creates constraints between camera poses. The link between pose 2 and 4 is based on the relative alignment of two 3D submaps. A relative pose constraint can be used between two multibeam submaps or between a multibeam submap and stereo camera based reconstruction.

matches. However, the algorithm presented here is more tractable in post processing given the large number of feature points used and the computing hardware available. Therefore the entire dead reckoned trajectory is used at once to find links. A link is hypothesized between two poses if they appear near each other in the dead reckoned navigation solution. The longer the survey, the more drift accumulates, increasing the covariance of the vehicle position and requiring a larger search radius. The size of this search radius grows unbounded with the length of the survey.

Stereo data association & bundle adjustment (Phase I)

The first factor graph is set up as a bundle adjustment problem [11]. It incorporates the odometry constraints with feature based constraints between stereo image pairs. The graph is solved to obtain the vehicle positions and the camera offset.

There are a number of ways to use images to constrain robot trajectories. Hover et al uses a 5DOF pose based image constraint [8]. Kunz uses a landmark based reprojection error minimization with a single camera [9]. Here however, a stereo system is available. A calibrated stereo system allows for a 6 DOF motion constraint unlike a monocular system, which can only provide 5 DOF constraints on motion due to loss of scale.

Links using stereo imagery are based on sparse feature point matching. First SIFT image features are extracted from the stereo image pairs in the link hypothesis. SIFT is essentially illumination invariant and requires little prepossessing for successful matching [13]. If images are particularly low contrast or have very uneven lighting, adaptive histogram equalization can be used to create more uniform feature extraction across the images. For each stereo pair, features are matched with each other. Matches that are more than five pixels from the epipolar line of their conjugate feature are rejected as poor matches. Typically thousands of stereo features can by matched at this step. SIFT descriptors in the left image of each

view are retained (Fig. 9).

Once stereo matching has been done on all image pairs, links between pairs are hypothesized and tested. The descriptors retained in the previous step are matched with the left image features in all the hypothesized link poses. Outlier feature matches between stereo poses need to be rejected so that they don't corrupt the factor graph solution. The matched features for each stereo pair are triangulated. Then a rigid motion model is fit to the triangulated features of hypothesized links using Least Median of Squares. Links are rejected if there are fewer than 6 matching features which fit the rigid motion model(Fig. 9). Though Figure 8 shows only two features which have been viewed by both pose 1 and 3, usually tens of features can be matched between the stereo pairs of two poses. The resulting measurement of stereo link generation is the image frame coordinates (u, v) of the matching features in the left and right images of each stereo pair. These points can be triangulated to create full 3D landmarks as viewed from each pose.

Camera and multibeam sensor offsets on the vehicle must be well aligned relative to each other so that their measurements can be properly aligned. These offsets are measured by hand and it is difficult to achieve the required precision. To avoid the guesswork, Kunz added an additional variable node to the graph: the camera offset node (Fig. 10). This additional variable accounts for the constant transform between the vehicle and sensor coordinates. The initial hand measurement of the camera offset serves as a prior on the offset node and the covariance of the prior encodes how well the offset was measured. It is worth mentioning that this node is often poorly constrained in the z direction and tends to float vertically. The vehicle is very stable in the pitch and roll directions which is the motion necessary to constrain z. This could make the map more difficult to geo-reference but has little impact on its self-consistency. The prior on the camera offset it the final constraint needed in the assembly of the bundle adjustment factor graph before it can be solved.

The location of the 3D landmarks projected on their respective images, along with the camera offset prior and the resampled navigation data are assembled into a graph. When the graph is solved, the result is a refined vehicle position at the time of every mapping sensor measurement, as well as an estimated camera offset.

Adding multibeam and cross modality constraints (Phase II)

After camera constraints have been used to refine the navigation and camera offsets, the data can be used to establish multibeam links. While these links have less influence on the over all navigation solution than the camera constraints, they are important for constraining the multibeam sensor offset ensuring proper alignment with the camera.

- **Submap assembly**

  The multibeam submaps are constructed by aligning adjacent pings in the submap coordinate system, s, using navigation data. First, the origin of the submap reference frame is assigned as the pose of the first multibeam ping of the submap. The individual multibeam pings are localized in s using the vehicle trajectory from Phase I. This data is segmented into submaps during the initial resampling phase. Navigation data is filtered according to section 2.3.3 and the multibeam pings are grouped into submaps. The submaps are ended when the covariance of the vehicle position relative to the submap origin reaches a certain threshold. The multibeam pings in a given submap are transformed into the submap coordinate system and each submap is considered a rigid point cloud.

  First, the origin of the submap reference frame is assigned as the pose of the first multibeam ping of the submap. The individual multibeam pings are localized in s using the vehicle trajectory from Phase I. This data is segmented into submaps during the initial resampling phase. Navigation data is filtered

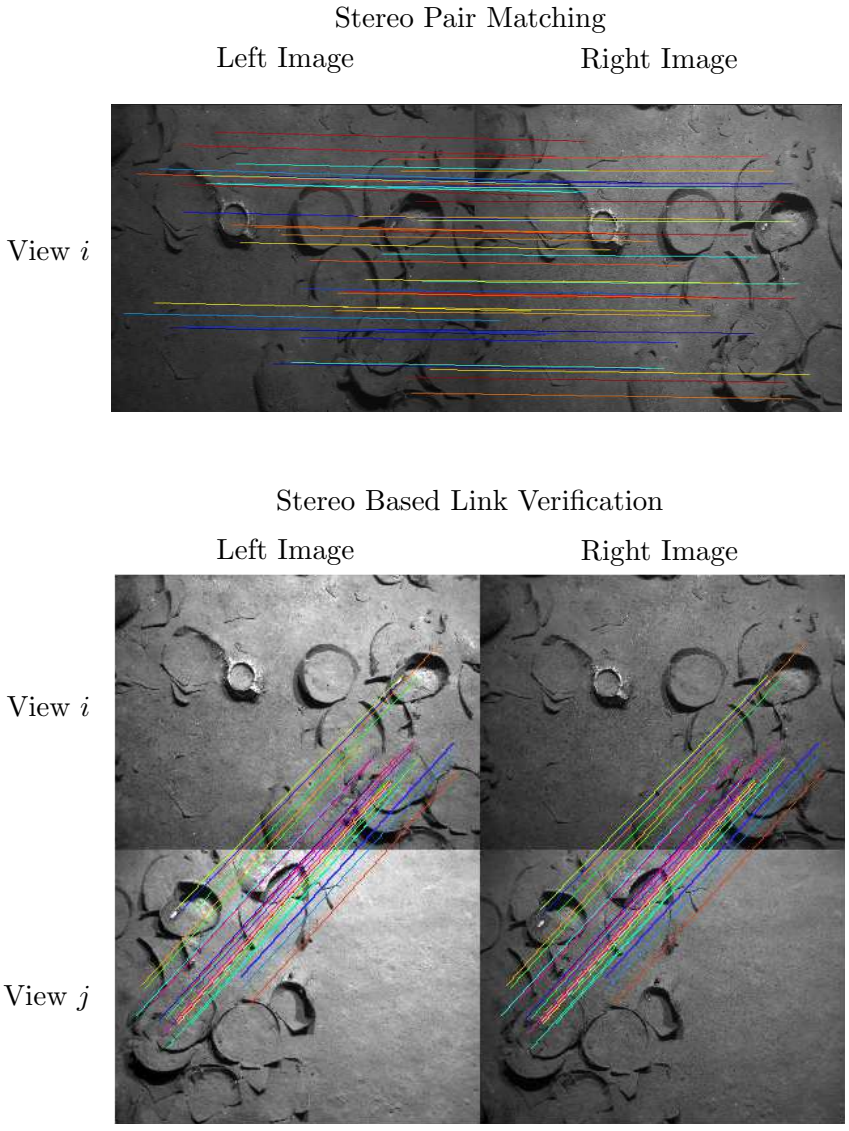


Figure 9. A verified link between stereo pairs. First stereo matching finds unique features which exist in both images of a stereo pair (top). Then these features are matched with similar features viewed in overlapping stereo pairs (bottom). This link provides spatial constraint on their relative positions of camera viewpoints at locations i and j.

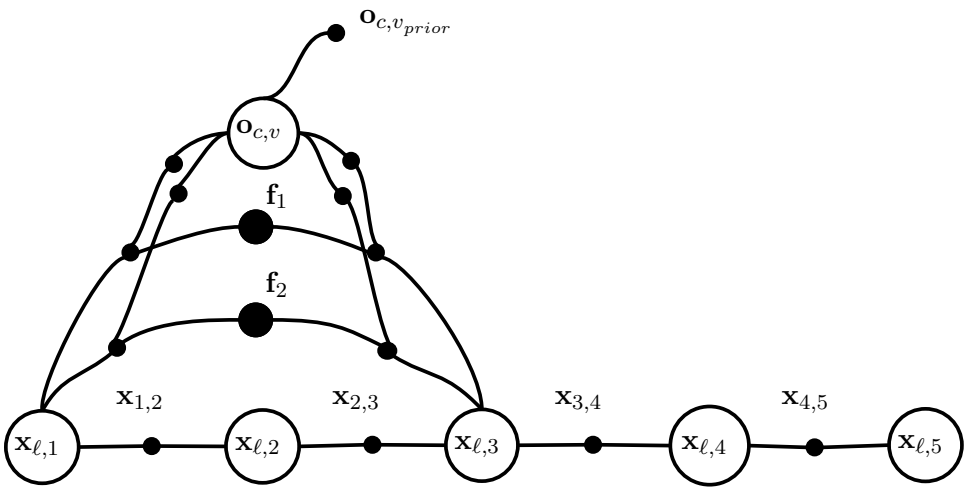


Figure 10. Bundle adjustment factor graph with camera offset node ($\mathbf{o}_{c,v}$). The offset node for the left camera is estimated concurrently with the vehicle poses at the time of each mapping sensor measurement.

according to section 2.3.3 and during this process, the multibeam pings are grouped into submaps. The submap is ended which the covariance of the vehicle position relative to the submap origin reaches a certain threshold. The multibeam pings in a given submap are transformed into the submap coordinate system and each submap is considered a rigid point cloud. The newly refined trajectory makes it unnecessary to break submaps according to accumulated error because the accumulated error has been corrected by the visual constraints. However, this process to break submaps is still used because it creates reasonably sized submaps in the case where there are very few or no successful imaging constraints.

These three dimensional submaps can be aligned with overlapping submaps to produce constraints on the relative position of the vehicle. Any relative pose constraints formed between submaps act on the vehicle pose which serves as the submap origin. This process is described in detail in [10] and refined for factor graph applications in [7].

- **Submap link alignment and verification**

Sonar submaps are aligned to constrain adjacent vehicle poses. In general, a relative pose constraint up to 6 DOFs found by aligning the submaps in x, y, z, roll, pitch and heading and computing the rigid transformation between their origins using point cloud registration techniques.

To establish link hypotheses, the submap boundaries are plotted in ℓ using the bundle adjusted vehicle navigation and the hand measured sonar offset. First link hypotheses are generated between potentially overlapping submaps then the overlapping regions are gridded (Fig. 11). The gridded data is aligned in the x and y directions by minimizing the Some of Squared Differences (SSD).

$$\Delta x, \Delta y = \underset{\Delta x, \Delta y}{min} \frac{1}{\|S_{\Delta x, \Delta y}\|} \sum_{x,y \in S_{\Delta x, \Delta y}} \left(z_{i,x,y} - z_{j,x+\Delta x, y+\Delta y}\right)^2 . \tag{1}$$

where $z_{i,x,y}$ is the depth of grid cell x, y in submap i, and $S_{\Delta x, \Delta y}$ is the set of all indices x, y is in submap i and $x + \Delta x, y + \Delta y$ is in submap j [10]. The minimum gives can be used to correct the initial estimate for the x and y components of the relative pose transform. If correlation is successful, a full 3D alignment is attempted with the SSD based alignment as an initial guess.

Point cloud registration has been widely researched for applications in robotics and scene reconstruction. Iterative Closest Point (ICP) in particular has become a common way to bring two point clouds into alignment by computing the rigid transformation between them [14]. ICP works by taking a random sample of points from one cloud, finding their nearest match in the other cloud and the computing the transform which pulls these points into the best alignment. This processes is iterated over for a pre-specified number of iterations.

ICP gives a full 6 DOF alignment between point clouds, however it is susceptible to local minima and sometimes converges to the wrong answer. After alignment, the ICP result is assessed to make sure it actually produces an

alignment improvement when compared with the SSD. Each point in one submap is linked to the nearest point in the other submap to determine the point-to-point error. This is done for the SSD as well as the ICP alignment results. The error histograms are summed and if the ICP error is mainly higher than that of the SSD, the ICP transform is rejected in favor of the 2 DOF SSD result. This method was developed by Roman [2].

The error surface of the SSD function is useful for determining the uncertainty of the link and for link rejection. A quadratic surface is fitted to the region around the minimum. The Hessian of this quadratic is the matrix of information gain for the link [9]. A large Hessian determinant indicates a very steep quadratic and a good minimum and large information gain from the link. Links are rejected if the determinant is less than 0.001. This value is only sensitive to the size of the region approximated by the quadratic. For the size of overlap and swatch width used in this thesis can be reasonably set to a 0.4m radius. If a larger region is approximated, the quadratic will tend to be not as steep even for good alignments, therefore the Hessian threshold has to be lowered.

When maps are only aligned in x and y, the factor node only constrains the graph in 2DOF. The link is given essentially zero information gain for all of the unconstrained degrees of freedom. In this case, the diagonal Hessian components are the information gain for x and y. When the ICP alignment is used, the Hessian is also used for x and y information gain and non-zero values are found empirically for the remaining degrees of freedom since the method in Roman, 2007 was found to underestimate information gain for these links to the point where the have no influence on the graph [15]. Instead the x and y information gain was taken from the 2 DOF information gain. Roll roll pitch and heading information gain was gradually increased until resulting map was at its most consistent and the links had moderate effect on the

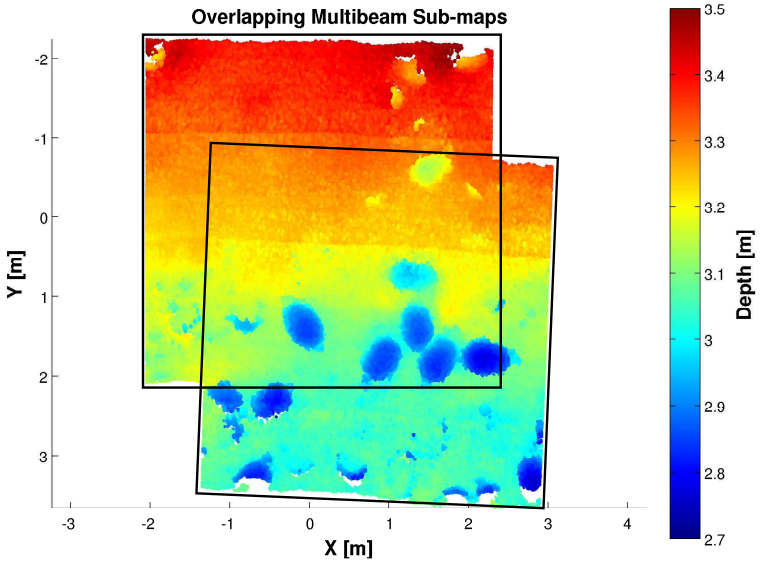


Figure 11. Multibeam submap alignment. Submaps are aligned by minimizing the SSD between the maps and the alignment is further refined if possibly using ICP point cloud registration.

multibeam sensor offset.

These links are also used to constrain the multibeam sensor offset. Unlike the camera offset, this offset can only be estimated in x, y, z and roll. Changing the offset in pitch or heading would change the shape of the submap which we assume is rigid. For a 6 DOF offset estimation to be valid, the submap would have to mutable and be re-aggregated for each new iteration of the navigation solver. Therefore, the offset was not allowed to vary in pitch and roll as the graph was optimized.

- **Cross modality links**

  Aligning the point clouds from the two sensors is critical to making a multi-modal map. To accomplish this, another constraint is introduced into the graph. This constraint connects the multibeam data to the camera data via a relative pose constraint between two respective submap origins. This is similar to multibeam data association(Fig. 12).

  The first step is to create stereo based submaps. This is done by perform-

ing dense stereo matching using the Block Matching technique. The stereo matches are triangulated to create a high point density reconstruction of the sea floor.

After the initial bundle adjustment and camera offset optimization, the resulting poses are used to select stereo submaps and overlapping multibeam submaps. Then link hypotheses are drawn between overlapping stereo and multibeam submaps. Currently this is done by hand-selecting a single multibeam submap and camera submap which overlap completely and contain significant structure to constrain alignment. The stereo images for the submaps are generally taken very close to the time associated with the multibeam submap origin. The relative positions between pose corresponding to these submaps is found by aligning the submaps using one of two methods. This relative pose is added as a constraint on the graph. Two types of constraint are investigated here to create the cross-modality constraint, one imposes a vertical constraint on the sensor offsets. The second aligns camera and multibeam subamps using full 6 DOF point cloud registration.

The first alignment method addresses the issue of poor constrained sensors offsets in the z direction. The two sensor offsets will tend to wander independently in the z direction when there is no constraint between them. A constraint which prevents this is required to keep the measurements of the two sensors mutually consistent. The relative position between the vehicle poses associated with these submaps was found by computing the average vertical distance between submaps. This distance was added as a 1 DOF vertical constraint between the poses attached to the submaps.

Another way to apply such a constraint is to use a 6 DOF constraint much like the one used to link two multibeam submaps. For cross modality links, a set of 15 camera and multibeam links containing reasonable amounts of structure were selected as link hypotheses. Then with the initial alignment

provided by the bundle adjusted navigation solution, the sum of squared differences was used to refine the alignment in x and y directions. Then ICP is performed to ascertain the full relative pose constraint between the camera and the multibeam submaps. This relative pose constraint is used to enforce the mutual alignment of the camera and multibeam point clouds.

For instance, say that an stereo pair is acquired at time i and a multibeam origin corresponds to time j. The cross modality link between vehicle pose at i and j is written as $\mathbf{x}_{i,j}$. This relative pose measurement between the two vehicle poses is the constraint which will be applied to the graph. It is a function of the relative pose between submap origins ($\mathbf{x}_{c_i,m_j}$), found using point cloud registration, and the sensor offsets:

$$\mathbf{x}_{i,j_{measured}} = \ominus \mathbf{o}_{c,v} \oplus (\mathbf{x}_{c_i,m_j} \oplus \mathbf{o}_{m,v}). \tag{2}$$

Navigation data between two poses close together in time has a very low covariance because there has been little opportunity for drift. Therefore and cross modality constraint between those two poses will tend to have more impact on the sensor offsets than they do on the navigation data. This prevents the multibeam sensor offset from floating away from the fixed camera offset when Phase II is solved. The Phase II graph containing the camera constraints, navigation constraints, multibeam constraints, multibeam offset prior, and cross modality constraints is solved to finally estimate the vehicle poses and multibeam sensor offset (Fig. 12).

2.3.4 Factor nodes: Error functions

The graph solution is inferred using a non-linear least squares solver to minimize the sum of the squared errors. The errors are computed from the error functions defined for each type of factor node. These measurement or error functions compute the error between the actual measurements and the measurements

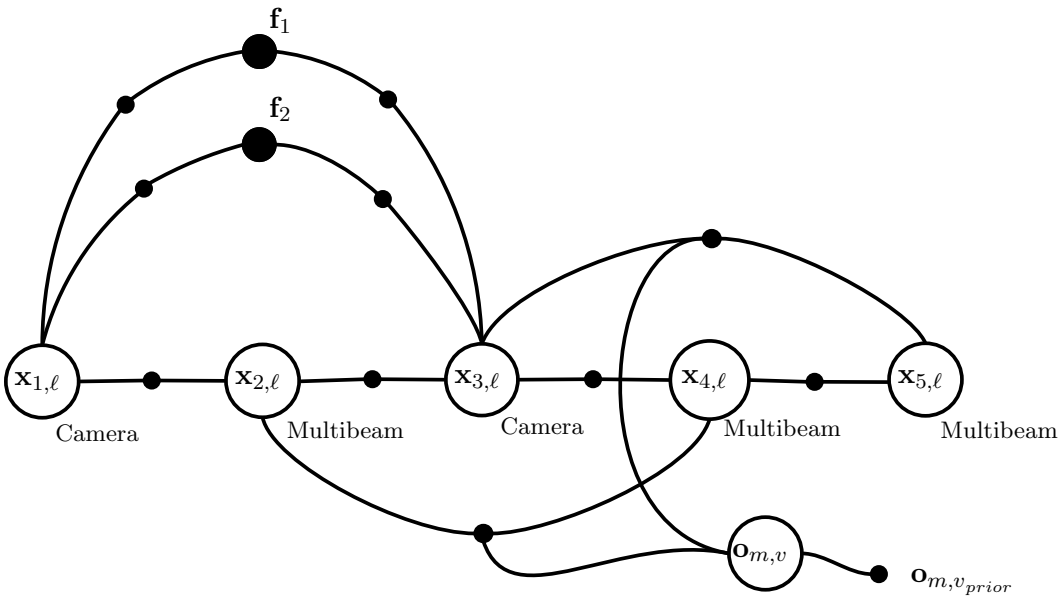


Figure 12. Multi-modal factor graph constraint enforces alignment between measurements from both camera and multibeam and simultaneously refines multibeam offset.

induced by the most recent estimate of each of the variable nodes.

Relative pose error functions

The relative pose error functions for multibeam-to-multibeam submap alignments and camera-to-multibeam submap alignments is straightforward. The error is defined as the difference between relative pose measured during the submap alignment step, and the relative pose induced by the most recent set of pose estimates (predicted relative pose).

The most general case allows for a 6 DOF relative pose constraint where $\mathbf{x}_{i,j} = [x, y, z, \theta, \phi, \psi]^\mathsf{T}$ is the full relative pose but, if only 2 DOF are constrained by the measurement, such as for multibeam submaps when ICP fails, the error is only computed with $\mathbf{x}_{i,j} = [x, y]^\mathsf{T}$. Here $\hat{\mathbf{x}}$ refers to the most current estimate of the relative pose vector and $\mathbf{r}$ is the error vector

$$\hat{\mathbf{x}}_{i,j_{predicted}} = \ominus \left(\hat{\mathbf{o}}_{m,v} \oplus \hat{\mathbf{x}}_{\ell,i}\right) \oplus \left(\hat{\mathbf{o}}_{m,v} \oplus \hat{\mathbf{x}}_{\ell,i}\right) \tag{3}$$

$$\mathbf{r} = \mathbf{x}_{i,j_{measured}} - \mathbf{x}_{i,j_{predicted}}. \tag{4}$$

Reprojection error function

The error between two stereo poses is computed using reprojection error. Reprojection error is a metric to simultaneously evaluate the correctness of camera poses and scene reconstruction. This is done by comparing the location of a feature based on the reprojection induced by estimated pose and scene to the same feature's actual location in an image (Fig. 13). Here $\mathtt{K}$ is the camera matrix for the left camera, image point $\mathbf{U} = [u, v]^\mathsf{T}$ and the 3D point in the camera reference frame $\mathbf{f}_c = [X_c, Y_c, Z_c]^\mathsf{T}$.

$$\mathbf{U}_{predicted} = \begin{bmatrix} (\mathtt{K}_{1,1}X_c + \mathtt{K}_{1,3}Z_c)/Z_c \\ (\mathtt{K}_{2,2}Y_c + \mathtt{K}_{2,3}Z_c)/Z_c \end{bmatrix} \tag{5}$$

$$\mathbf{r} = \mathbf{U}_{measured} - \mathbf{U}_{predicted} \tag{6}$$

$\mathbf{f}_c = [X, Y, Z]^\mathsf{T}$ can be expressed in the camera coordinate frame by

$$\mathbf{f}_v = {}^{\ell}_{v}\mathtt{R}\hat{\mathbf{f}}_\ell + {}^{v}\mathbf{t}_{\ell v}$$

$$\mathbf{f}_c = {}^{v}_{c}\mathtt{R}\hat{\mathbf{f}}_v + {}^{c}\mathbf{t}_{v,c}$$

From the reprojection error equation, it is reprojection error can be evaluated using only one camera at each vehicle pose. Stereo image pairs are useful however, for several reasons. First of all, having two cameras allows a point to be triangulated which gives a good 3d initialization in ℓ. For monocular vision, the point depth is unconstrained in distance along the ray passing through the camera focal point and the image feature. An estimate of this distance is often approximate, perhaps set to the vehicle altitude at the time of image capture. Less precise initial estimates induce weaker constraints on the graph.

The constraints between image poses are enforced by minimizing reprojection error over the associated poses and landmarks. Reprojection error is determined by comparing the position of an object in the image to the position of the actual object backprojected onto the image using the camera matrix, camera pose, and object pose. The Euclidean distance between the backprojected object and its

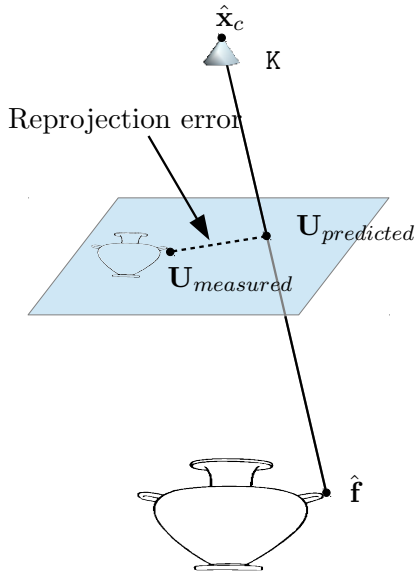


Figure 13. Reprojection error, r is a metric for evaluating the accuracy of the estimated camera pose ($\hat{\mathbf{x}}_c$) with respect to the estimated feature position ($\hat{\mathbf{f}}$).

image is the reprojection error. The graph solver however minimizes over a basic error vector containing the error $\mathbf{r} = [u_{error}, v_{error}]^\mathsf{T}$.

2.3.5 Error metrics

The quality of the map assembled from optimized navigation can be evaluated using several types of error metric. The first type are the error metrics over which the graph was optimized. The second type are error metrics which arise from constructing a multi-modal map and evaluating its characteristics directly. It is important to distinguish between these two types of error metrics.

Ideally the pose graph would be solved by minimizing an error metric which best expresses map quality. This might be an error metric which expresses the alignment of the submap point clouds. Unfortunately, such a function does not have very well defined local minimum and would have a hard time converging. Instead we optimize over more constrained error functions which have clear minima are good approximations for overall map alignment. Ultimately however, it is vital to know how well the point clouds align since this is a good predictor of final map

quality.

Optimized error metrics

There are two distinct error metrics which are optimized over during graph inference. The first is the residual of the relative pose estimation given in Equation 4. The next is the reprojection error calculation shown in Equation 6.

These two residuals are useful for evaluating the quality of graph inference. They can give insight into potential outliers and assist in finding problems in preliminary processing. Areas of the graph which contain relatively large residuals might contain a bad links indicating the need for robust inference methods, or some other error. However, these methods don't give very much information about the quality of the map that might be constructed from the optimized navigation data.

Reprojection error computed over the estimated position of the features gives some indication of how consistent the stereo point cloud is with the images. Relative pose residuals evaluate how well submaps alignments were enforced in the final navigation solution but do not directly evaluate how well camera submaps align in the final map.

The assumption at this point is that consistent poses should lead to consistent point cloud alignments. However, since all the constraints are not directly based on point cloud alignment, and instead reduce point cloud alignment to lower dimensional approximation such reprojection error and relative pose error, maximizing this reduced approximation of point cloud consistency does not necessarily lead to more consistent maps.

Map based error metrics

The error metrics for reprojection error and relative pose constraints are practical approximations for 3D structure alignment, which are proxies for map alignment error. Since this is the case, its is valuable to examine map quality directly to ensure that it is sufficiently improved by navigation refinement proceed with

the mapping steps.

- **Map alignment**

  A composite map is composed of a number of camera and multibeam submaps projected into a common coordinate frame. This error metric concerns the quality of submap alignment across the entire composite map. The error metric is based on the Map-to-Map error developed by Roman [2], that has been modified to accommodate comparisons between submaps acquired by different modalities.

  The metric is derived from the idea of the Hausdorff distance [16]. It quantifies error between multiple submap point clouds and assigns an error value to each cell of the gridded composite map. This particular implementation is designed to evaluate the distance between submaps produced by different sensing modalities which may have different sampling densities. The implementation works as follows: A grid is laid out in ℓ and the composite point cloud containing all of the submaps is projected onto it. Points are assigned to the cells that they are projected into and labeled with their submap number. A point $\mathbf{X}_i$ from map $\mathcal{M}_i$ is selected at random where i is all of the maps present in the grid cell. A plane $\mathbf{p}_j$ is fit to the points representing each map $\mathcal{M}_j$ in that cell and the adjacent cells. The distance (d_{ij}) from X_i to $\mathbf{p}_j$ is computed for each value of j and the maximum value of d_{ij} is noted. Multiple points X_i can be selected to produce multiple d_{ij} and the average is taken. This process is repeated $\forall i \in \mathcal{M}$ and the mean d_{ij} is taken to be the map-to-map error for that cell (Fig. 14).

  In the previous implementation, a plane was not fit to the set of points in $\mathcal{M}_j$, instead the distance between $\mathbf{X}_i$ and the nearest point in $\mathcal{M}_j$ was used. This is a reasonable approach when sampling densities are consistent and greater than grid cell size. submaps made from stereo cameras in particular

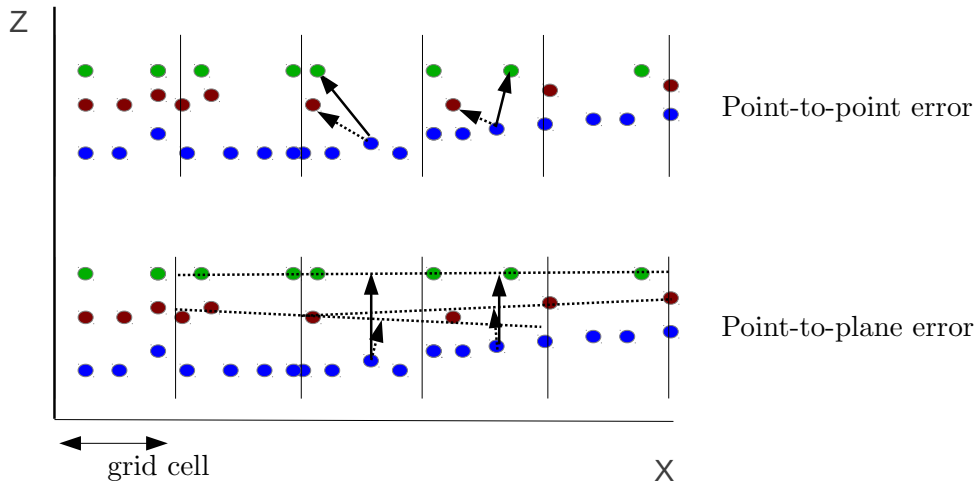


Figure 14. Point-to-point versus point-to-plane error. For point-to-point the distance between a point in one map and its nearest point in each of the other maps is computed, then the longest distance is retained as error. This can artificially inflate the error in cases of irregular sampling. Point-to-plane uses the distance the point in one map and a plane fit to the local area of the other maps. This reduces the impact of irregular sampling of the surface.

are subject to inconsistent sampling. Dense stereo methods often result in irregular spacing as well dependent on the photometric characteristics of the images. In order to properly capture the distance between point clouds, without over inflating it, it is important to use point to plane in stead of point-to-point error.

When this cell by cell error metric is evaluated across all of the submaps (including stereo and multibeam) , the result is a gridded representation of the map-to-map error. This is a good illustration of the quality of alignment between the submaps which will ultimately comprise the final composite map of the surface.

Another way to use this metric is to assign all multibeam submaps to one submap number and all camera submaps to another map label. Evaluating the map to map error over these two 'submaps' gives a sense for how well the two modalities are aligned. This is an important thing to examine since it is

well established that the individual modalities can form self consistent maps, but no one has ever investigated their how consistent the are with each other.

- **Texture alignment**

  Texture alignment refers to how well we can expect images projected onto the map structure to line up with each other and it can be evaluated using reprojection error. 3D locations of features appears in two camera poses are backprojected into the opposing viewpoint and the backprojected point is compared to the known feature location in that image to get reprojection error. In the previous section, this error was evaluated using the feature locations optimized during navigation refinement, however this is not truly reflective of the alignment of texture maps when projected on the mesh since the texture maps are not warped to match the unrefined feature locations. Instead textures will map to the mesh of the unrefined feature locations . Therefore, reprojection error will best evaluate texture alignment if done with the initial feature locations reprojected into the refined camera poses. While no texture mapping is done in this thesis, that would be a logical and straightforward way to extend the utility of the work. Reprojection error is also a useful approximation for how well the camera meshes align with each other.

2.4 Results

The results of navigation refinement dictate the quality of the ultimate map so it is important to understand the characteristics and breakdown points of this process. The various data association techniques contribute an important set of constraints to the navigation refinement solution. In particular, the use of cross modality registration has been introduced as a new constraint and the results presented here. Overall, the navigation refinement results can be evaluated in terms of the error metrics summarized in the previous section. These error metrics give an

indication of expected mapping performance as well as give insight into particular considerations which should be made in developing the mapping methods.

2.4.1 Data association

Links between poses constrain the vehicle to locations which will provide self consistent maps. This section summarizes the utility and breakdown points of each of the data association techniques used to constrain the vehicle poses.

Stereo

Stereo data association is based on two stereo poses viewing the same feature. Figure 15 shows verified links plotted on the refined navigation data. The lines join poses which have viewed the same feature. To ensure that outliers are rejected, only links between poses which share six or more features consistent with the same rigid motion are verified as links. While the links are evenly spread throughout most of the graph, there are relatively few links between the body of the survey and the diagonal crossing line.

Stereo links fail in poor imaging conditions. If light is poor, turbidity is high, scene texture is lacking, or the viewpoint between images is too different, there are several points where the algorithm will catch bad links:

1. If turbidity is high or the vehicle is too far from the bottom, few or no stereo matches will be made between images in the stereo pairs, thus no SIFT descriptors will be available for matching with hypothesized links pairs.

2. If the scene has changed between one measurement and the next due to silt kick up or moving fish, few common SIFT descriptors will be found between hypothesized link image pairs. If any are found, they may not be consistent with a 6 DOF rigid motion so the features matches between stereo pairs will all be rejected and no link will be verified.

One portion of many surveys where stereo links tend to fail is along diagonal

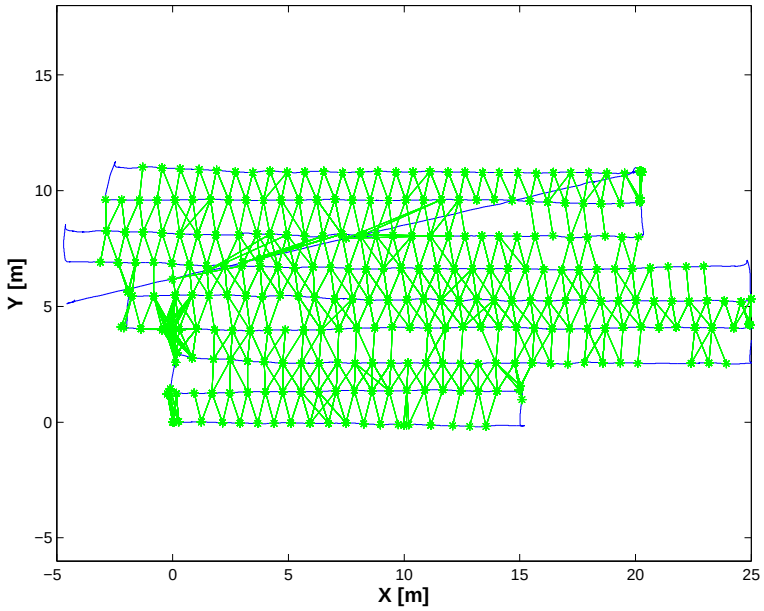
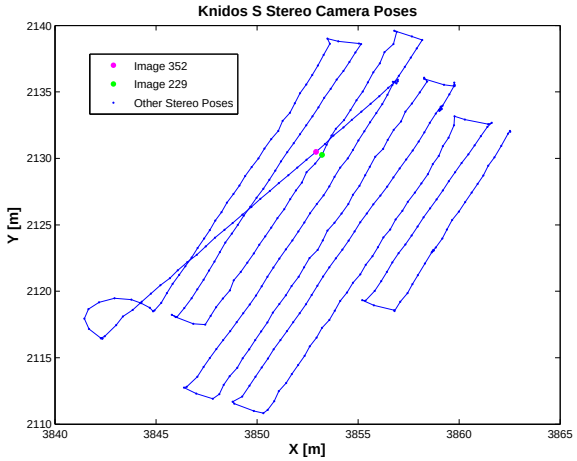


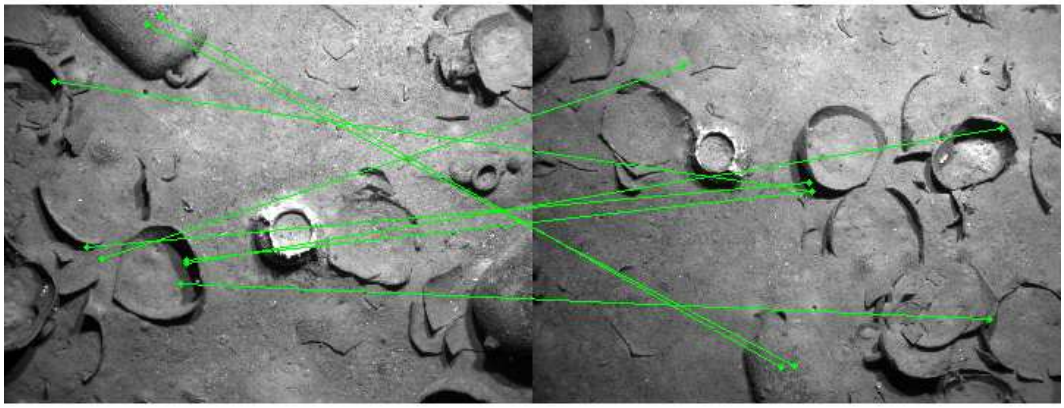
Figure 15. Verified stereo links. The links shown here represent poses which share a view of six or more features, each feature is one link between those poses.

loop closures. Along the diagonal loop closure shown in Figure 16(a), the pink point and green point corresponding to stereo pairs 352 and 229 respectively are close together so a link is hypothesized. The left hand image from each pair is shown in Figure 16(b). The images contain enough distinct and common features that a link ought to be easily obtained. However, the 11 sift matches overlaid on the two images are incorrect except for two. There are not enough correct matches to meet the threshold so this link is rejected. Even so, six links were successfully established between the survey and the crossing line.

One way to asses bad links is using reprojection error. Any feature with a much higher final average than the others is likely to be an outlier. Figure 19 shows no reprojection errors which are inordinately high after optimization, meaning that there are probably no bad links and a navigation solution consistent with all of the linking features was obtained.



(a) Stereo Camera Poses



(b) Feature Matches between left images of 229 and 352

Figure 16. **(a)** The pink and green points indicate the pose of the stereo rig for image pairs 353 and 229 respectively. While the poses are close together and the images overlap, no link was formed here because the scene was lit from a different angle for each pair and link verification matching was unsuccessful. The only matches that were found are displayed in **(b)** and are clearly incorrect.

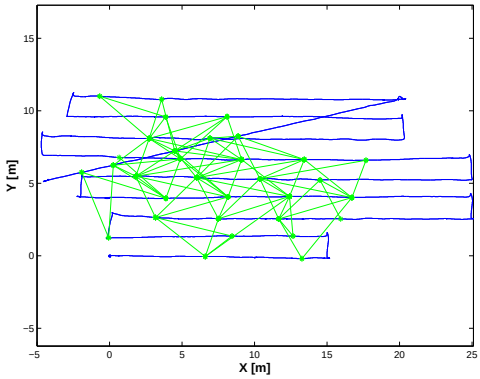


Figure 17. Submap links verified on bundle adjusted navigation data. This figure shows that the majority of the links are clustered near the center of the survey. This corresponds to areas where there is more structure.

Multibeam

Multibeam data association is executed after the bundle adjustment step. As a result, most of the drift has been removed from the navigation data. This gives good initial alignment for the multibeam relative pose estimates. Figure 17 shows the distribution of verified links established between multibeam submaps assembled using the bundle adjusted navigation data. There are 76 total links distributed throughout the survey.

Submap links are based on the alignment of scene structure, therefore if there is little scene structure, alignment is less likely to be successful. Figure 17 shows that there is more concentration of links in the center of the survey where there is more structure from the debris of the shipwreck. There aren't as many multibeam links as there are stereo links, and they appear to not alter the navigation data very much from the bundle adjusted solution, however they are useful because they enforce self consistency between multibeam submaps as the cross modality links enforce consistency between multibeam submaps and camera submaps.

Cross Modality registration

The cross modality registration uses stereo and multibeam sonar range data and aligns them. Two different methods with different levels of constraint were

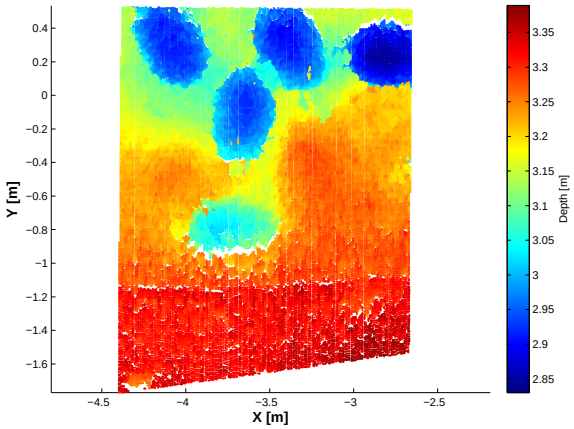
used. The first method constraining only the z direction between aligned submaps is the only option when there is no scene structure. The second uses the well known ICP point cloud registration algorithm to compute 6 DOF constraint between the two overlapping submaps. The results of ICP alignment process are presented here, and the impact that both methods have on the navigation solution is demonstrated in Section 2.4.2.

The ICP alignment with selected submaps had a good success rate. 15 multibeam-stereo pairs that were selected. SSD gave a distinct results for 12 of them and all 12 converged consistently to reasonable solutions. It was helpful to select submaps covering areas with structure. An advantage of using dense stereo reconstructions is the very high point density available which provides flexibility. All of the stereo points could be used, but at a drastically increased processing time. Instead camera reconstructions were down-sampled to 1.5 points/cm^2 density to match the multibeam's natural point density of 1.5 points/cm^2. It was also useful to remove outliers from the dense stereo by gridding both point clouds and removing stereo data more than three standard deviations from the mean. ICP convergence generally occurred between 4 and 10 iterations.

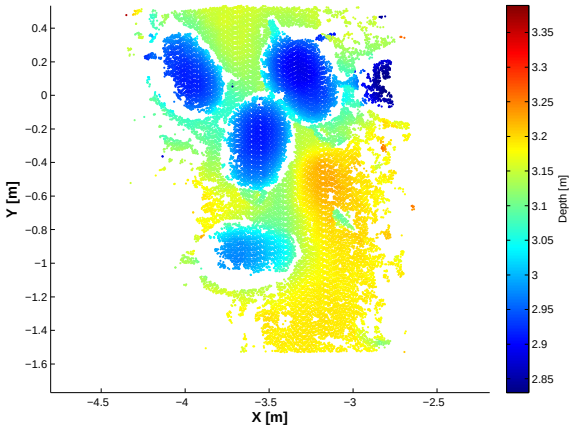
Figure 18 shows the typical results of aligning camera and multibeam submaps from the area shown in 18(a). Final alignments showed very little error when evaluated using the map-to-map error metric (Fig. 18(e)), however there are gaps around the edge of the objects due to occlusions. These gaps do not prevent ICP from converging but they contribute ambiguity to the alignment.

ICP has been a successful method for registering stereo submaps to multibeam submaps because these submaps have achieve the required sampling density, and have good alignment. Dense stereo techniques allow flexibility in selecting sampling density which results in convergence of the ICP algorithm. Additionally, since the navigation data has already gone through one round of refinement, the initial alignment between submaps is good.

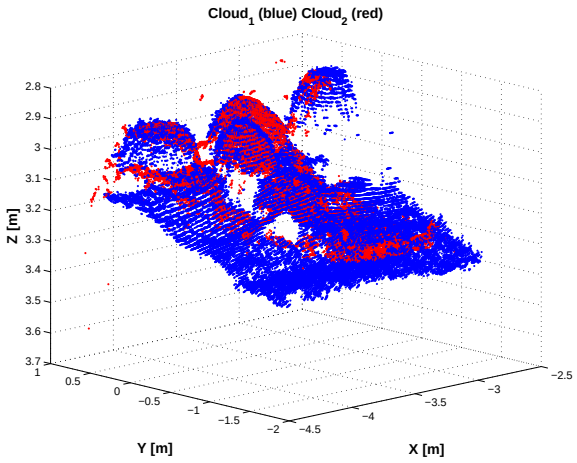
(a) Image of Region



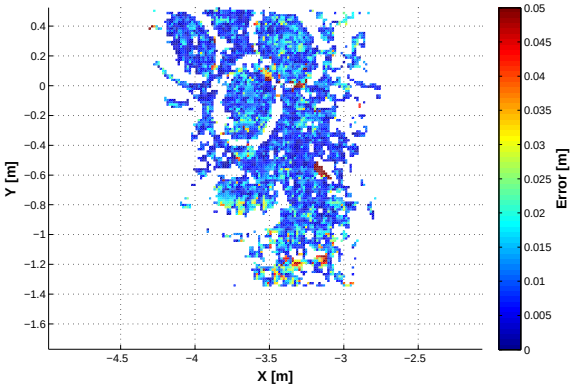
(b) Multibeam Data



(c) Camera Data



(d) Aligned Point Clouds



(e) Error Map

Figure 18. Cross modality registration. Two sections of multibeam **(b)** and stereo reconstruction **(c)** are aligned using point cloud registration techniques **(d)**. The quality of the alignment can be assessed using the map-to-map error metric **(e)**.

2.4.2 Factor graph results evaluated using error metrics

The output of the factor graph inference and the impacts of the various constraints are evaluated using the error metrics outlined previously. These error metrics reveal information about the improvement in the submap alignment as well as remaining artifacts. In addition they illustrate the utility of cross modality links.

Reprojection error

The navigation solution is computed by minimizing error over a number of functions, one of which is reprojection error. Reprojection error is a good metric for comparing various navigation solutions because it reflects approximately how well images will line up when they are projected on the 3D map structure. Kunz shows that adding camera constraints and camera offset estimation improves reprojection error [9]. Additionally it was shown that multibeam relative pose constraints do not worsen the reprojection error and those results have been reproduced here. Figure 19 shows that the addition of cross modality links also does not negatively impact the reprojection error of the solution. Next it will be shown that cross modality links also improve the mutual consistency of data from the two sensors.

Map-to-map error

To evaluate mutual alignment of the two sensors, map-to-map error is used. The dense stereo reconstructions are counted as one map and the multibeam submaps are counted as another. To evaluate the overall error characteristics of the map, each submap is treated separately in the map-to-map error calculation.

First map-to-map error is used to show the impact of cross modality links on the alignment between the two sensors. A histogram of map-to-map errors is useful when the amount of error is great enough that a spatial distribution plot becomes difficult to interpret. In this histogram the error for no cross modality links is large (Fig. 20). Adding either z links or ICP links substantially reduces

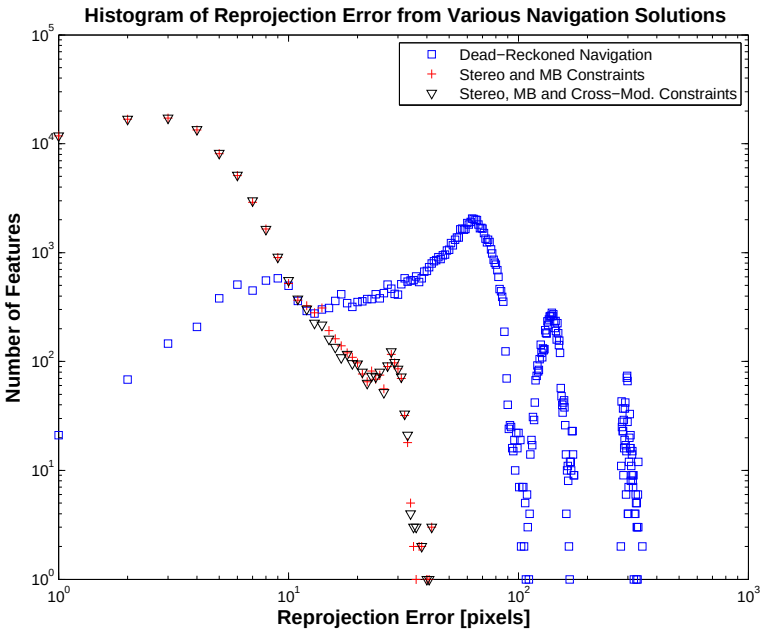


Figure 19. Reprojection error results. Adding cross modality constraints doesn't degrade the reprojection error.

this error. Note that the distribution between the a z links and the ICP links are quite similar.

Examining the spatial distribution of error will give some indication of whether these cross links have an impact on the alignment of the two sensors in x and y. In fact, it appears that between the 1 DOF and the 6 DOF alignment, there is very little difference in the spatial distribution of error (Fig. 21).

If each dense stereo reconstruction and multibeam submap is labeled as a different map and then map-to-map error is computed, it is an indicator of overall point cloud thickness (Fig. 22). The most obvious error is at the edges of objects where slight misalignments between submaps are apparent and error is often as big the object is tall. The error appears at the edge of every object in the map and tends to be a consistent width. This indicates either a constant bias in the relative offset between the two sensors, or that the sensors resolve edges differently than one another.

Another error type of error appears as a gradual increase in error across the

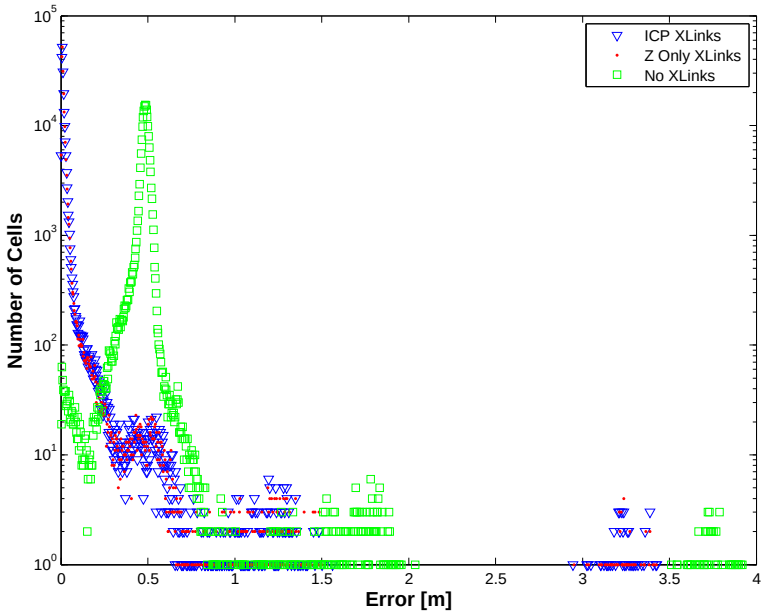


Figure 20. Histogram of map-to-map error. Adding cross links significantly reduces the alignment error between camera and multibeam maps.

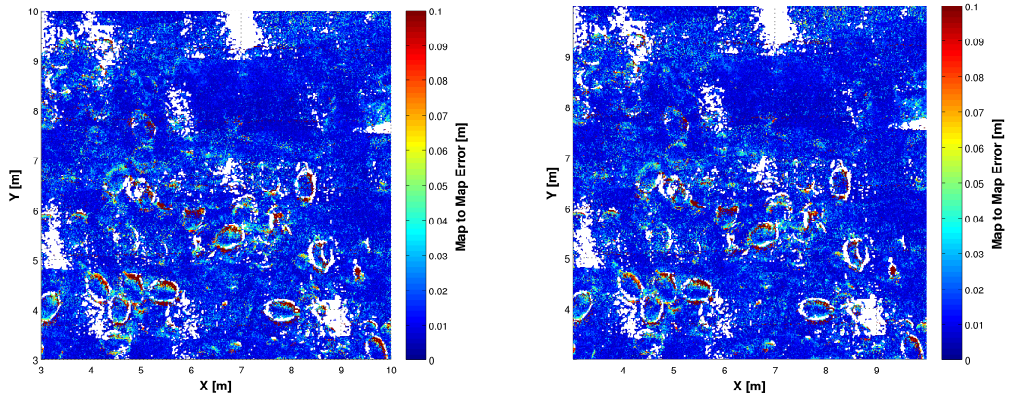
width of each trackline (Fig. 22). This is a slight roll bias in the camera offset. This type of error is usually apparent in the map as well and indicates that the offset wasn't fully corrected during the navigation refinement step.

2.5 Discussion

This chapter has focused on the necessary steps for aligning multibeam and stereo in the same coordinate system by refining navigation data. The motivation for this is to align the two modalities well enough that a single map can be constructed from the fused data sets. Additionally, several error metrics for evaluating the results have been reviewed.

2.5.1 Data association

One problem with the presented approach to link hypothesis generation for data association is that as maps get larger, more navigation drift occurs and the search radius for potential links must be wider. This adds computation time but this is not a large concern when the solutions are computed in post processing.



(a) Error with z only Cross Links

(b) Error with 6 DOF Cross Links

Figure 21. Closeup of map-to-map error with two different types of cross modality links. There is not any obvious difference in spatial distribution or error between 1 DOF versus 6 DOF cross modality constraints.

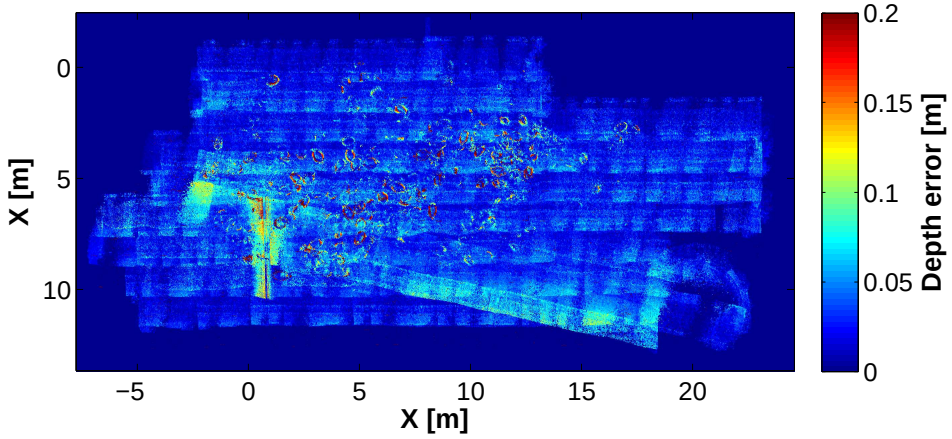


Figure 22. Map-to-map error with cross modality links. The errors shown here are indicative of the quality of submap alignment. The majority of map to map errors are at the edges of objects where slight misalignments in x and y produce errors in z equal to the height of the objects.

Another problem is that with so many links being compared, there is more room for bad matches. To cope with this, we use aggressive outlier rejection thresholds during link verification. This we can reliably reject bad links and have found that a large number of good links remain.

Stereo based data association

Matching stereo based measurements with each other to form links is a strong way to constrain the navigation data. It performs well, providing constraints on most images even in areas of the sea floor with few structural features since the textural composition of sea floor tends to be rich enough for unique feature matches.

The outlier rejection threshold requiring six matched features between poses is somewhat aggressive but it doesn't cause too many problems because good matches are so prolific. However, it may be possible to avoid this during the navigation solution by incorporating outlier rejection into the navigation solution and rejecting features which low marginal probability at each iteration. Another option may be to use a robust error function, though early experiments show that this approach often fails to converge to a solution.

In spite of the general success with stereo data association, there are much fewer links associated with images on the crossing line. This is effect is particularly evident in high relief scenes where lighting and parallax create different effects as viewpoint changes. There are ways to avoid this problem. First, the vehicle can close loops over texturally but not structurally rich areas where lighting and parallax will cause fewer differences between viewpoints, but available texture still provides substance for good data associations. Another option is to close loops with the vehicle at the same heading as was maintained during the survey to achieve similar lighting and projective characteristics. The problem here is that profiling sensors such as multibeam sonar or structured light require heading and course over ground to be the same for proper data acquisition and coverage, so

this approach isn't practical if a loop closure is necessary with those instruments. Finally, more distributed lighting systems (not feasible on *Hercules* due to space constraints) are begin used on other vehicles to mitigate the problem of shadows.

Cross modality links

Relative pose between camera and multibeam point clouds can be established effectively using point cloud registration techniques to create a cross modality link, however the characteristics of the individual sensors may impact the quality of this registration.

Camera and multibeam have different susceptibility to occlusions. Stereo cameras are unable to provide depth information for any area of the scene which isn't visible in two views. In scenes with large amounts of relief, occlusions become more obvious farther from the center of the stereo reconstruction (Fig. 18(c)). Farther from the center of the point cloud, there are more gaps in the data corresponding with occlusions. On the other hand, multibeam is somewhat less sensitive to occlusions. First of all, a point only has to be visible from one viewpoint, instead of the two required for stereo. Second, since it is a profiling instrument, occlusions only increase as a function of across track distance from the instrument center. There are no along track occlusions which are present in the stereo reconstructions. These occlusions are a limiting factor in the quality of the registration. That said, ICP accomplishes alignment for the places where there is data from two sensors (Fig. 18(e)). The amount of occlusion present implies a corresponding amount of uncertainty in the alignment between the sensors.

The process of generating link hypotheses between multiple modalities is currently done by hand. It would not be a stretch to automate, however. Link hypotheses could be generated based on areas where there is significant overlap between multibeam and stereo. To reduce the over all number of hypotheses and improve performance, a metric related to the normals of the surface could be used to determine which submaps have enough structure to be worth matching. If the

survey contains little structure z links can be used instead.

It is interesting that even with only z links, the alignment between the two sensors is good. This indicates that the navigation is constrained well enough that the submaps tend towards good x, y alignment even without cross modality constraint in those directions. However, the full cross modality links have a lot of value because they are a step towards aligning maps from two different sensors taken during two different surveys.

The necessity of hand tuning the information gain for cross modality links indicates that there is some unresolved issue in determining information gain for point cloud alignment which undervalues the link. Another possibility is the odometry or stereo constraints are being over valued. Resolving this issue requires further investigation since the relative importance of the constraints is important to the quality of the result.

2.5.2 Map-to-map error and implications for mapping results

It is important to know if the navigation is good enough for map construction. Figure 22 indicates that we can expect residual navigation error. This manifests at the edges of objects and tracklines. This same figure also shows $\sim$ 3cm thickness to the point cloud even where there are no objects. This point cloud thickness can be partially attributed to the natural variance in the range measurements. It is also related to sensor offset refinement an apparent roll error which corresponds to camera submaps.

Assuming that this navigation solution is the best available, the next chapter undertakes the goal of creating a map that combines the strengths of each of the sensors.

List of References

[1] R. Smith, M. Self, and P. Cheeseman, "Estimating uncertain spatial relationships in robotics," *Proceedings of the Second Annual Conference on Uncertainty in Artificial Intelligence*, pp. 167–193, 1986.

[2] C. N. Roman, "Self consistent bathymetric mapping from robotic vehicles in the deep ocean," Ph.D. dissertation, MIT/WHOI Joint Program, 2005.

[3] R. M. Eustice, "Large-area visually augmented navigation for autonomous underwater vehicles," Ph.D. dissertation, MIT/WHOI Joint Program, 2005.

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

[5] I. Mahon, S. Williams, O. Pizarro, and M. Johnson-Roberson, "Efficient view-based SLAM using visual loop closures," *Robotics, IEEE Transactions on*, vol. 24, no. 5, pp. 1002–1014, Oct. 2008.

[6] F. Dellaert and M. Kaess, "Square Root SAM: Simultaneous localization and mapping via square root information smoothing," *Intl. J. of Robotics Research (IJRR)*, vol. 25, no. 12, pp. 1181–1204, Dec 2006.

[7] I. Vaughn, "Microbathymetry using self-contained navigation and simultaneous localization and mapping," Master's thesis, University of Rhode Island, 2012.

[8] F. Hover, R. Eustice, A. Kim, B. Englot, H. Johannsson, M. Kaess, and J. Leonard, "Advanced perception, navigation and planning for autonomous in-water ship hull inspection," *Intl. J. of Robotics Research, IJRR*, vol. 31, no. 12, pp. 1445–1464, Oct 2012.

[9] C. Kunz, "Autonomous underwater vehicle navigation and mapping in dynamic, unstructured environments," Ph.D. dissertation, MIT-WHOI Joint Program, November 2011.

[10] C. Roman and H. Singh, "A Self-Consistent bathymetric mapping algorithm," *Journal of Field Robotics*, vol. 24, no. 1-2, pp. 23–50, 2007.

[11] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustmenta modern synthesis," in *Vision algorithms: theory and practice*. Springer, 2000, pp. 298–372.

[12] J. C. Kinsey and L. L. Whitcomb, "Preliminary field experience with the dvlnav integrated navigation system for manned and unmanned submersibles," In: Proceedings of the 1st IFAC Workshop on Guidance and Control of Underwater Vehicles, GCUV 03, Tech. Rep., 2003.

[13] D. Lowe, "Object recognition from scale invariant feature descriptors," *Computer Vision, IEEE Conference on*, p. 1150, 1999.

[14] P. Besl and N. McKay, "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[15] C. Roman and H. Singh, "Improved vehicle based multibeam bathymetry using sub-maps and slam," in *IROS'05: Proceedings of the 2005 IEEE/RSJ international conference on Intelligent robots and systems*, 2005, pp. 3662–3669.

[16] P. Cignoni, C. Rocchini, and R. Scopigno, "Metro: measuring error on simplified surfaces," in *Computer Graphics Forum*, vol. 17, no. 2. Wiley Online Library, 1998, pp. 167–174.

CHAPTER 3

Mapping

3.1 Introduction

This chapter focuses on producing a bathymetric map using data from two mapping sensors. It is assumed that the sensor poses have been established already during the navigation refinement step. In general, the fidelity of a map is evaluated by the end user using somewhat abstract characteristics related to the map's specific purpose. Several such characteristics of a useful map are distilled into concrete criteria. The ways that hybrid maps can address these criteria provide justification for producing them as an alternative to the current methodologies. Two possible methods for combining multi-modal 3D point cloud data will be presented. The first method serves as a basis for comparison and the second method addresses issues of multi-modal data fusion and outlier rejection with emphasis on different aspects of map fidelity. Finally, the resulting point cloud will be evaluated in terms of how well it addresses the map fidelity criteria.

3.2 Background

A number of methods have been developed for producing reconstructions of the sea floor from images and acoustics.

3.2.1 Photomosaics

Photomosaics are maps comprised of images which are registered and warped to bring them into alignment and blended together. This is a fairly simple problem for a few images but underwater surveys are often made up of hundreds to thousands of images [1]. If many images are warped and aligned naively, major distortions can occur. Resolving this type of error requires a global solution to determine the projective transformations for each individual image which distribute warping evenly across the map. It done properly no section of the map is subject

to more distortion than the others [2]. This approach to mapping produces flat maps which convey a large amount of information about shape and texture of the scene. Such mosaics have been a very useful for archeology since they provide information about the relative position of artifacts and allow scientists to visualize an entire underwater scene at once [3].

The underlying assumption of this type of mapping is that each image is of a planar scene. In practice this is regularly violated. The result is that the map is not a scale accurate representation of the scene. In spite of this photomosaicking is still widely used because well automated solutions are available which makes such maps easy to produce and ultimately, they very informative in spite of their drawbacks.

3.2.2 2.5D and 3D maps

The next level of complexity in mapping is creating a model which conveys shape in 2.5D or 3D. A wide variety of methods have been developed to accomplish this, using both acoustics and optics.

Typical approaches such as SFM have been employed using sparse features [4, 5]. In feature rich areas the resulting mesh is very accurate and quite dense. They also rely on sparse feature extraction, which can be tailored to focus on high relief areas and areas of geometric importance, so complicated terrain can be efficiently represented. However, since the sparse points are optimized during the structure from motion solution, this method does not lend itself to arbitrarily high point densities. It is ultimately limited by the feature extractor's ability to extract and match features and the computational burden to optimize feature locations.

Other approaches which use acoustic data such as CUBE and BP-SLAM build height maps using a Bayesian filter approach. Depth measurements are added to a graph or grid structure and redundant measurements are fused with a filter [6, 7]. These approaches are successful but have never been adapted for multiple modalities. One problem with applying them to multiple modalities is that naively

fusing two modalities together may reduce the quality of the more precise sensor.

The previously mentioned approaches (except CUBE) refine navigation data while building the map. Another option is mapping from known poses [8]. Roberson assumes known poses and reconstructs a surface using stereo vision data [9]. When a mesh is built from known feature points and poses, producing a seamless reconstruction becomes an issue of mesh and texture blending [10, 11]. This type of rendering produces very appealing maps with good local and global accuracy. Blending textures and meshes however can disguise alignment issues.

Mapping with known poses has some characteristics which are useful for multi-modal mapping, mainly the idea of splitting navigation and mapping into two different steps. However, blending two modalities together without taking into account the characteristics of each sensor may reduce the detail portrayed in the final surface reconstruction.

3.2.3 Multi-modal mapping

Multi-modal mapping requires specific considerations for the characteristics of each sensor. Previous attempts have been limited to computing scene structure with multibeam and overlaying texture with the images [12, 13]. However, stereo vision range data has some appealing characteristics that can compliment multibeam scene reconstruction. A next logical step in multi-modal sea floor mapping is to synthesize a sea floor reconstruction from both multibeam sonar and stereo vision.

Microbathymetric mapping at a scale of $\mathcal{O}$(5cm) surface reconstructions of the sea floor can benefit from merged data [14]. A final surface at this scale can be overlaid with image data from a camera to create map which conveys detailed shape and texture of the sea floor [9, 12, 13]

Such surface reconstructions can be thought of as 2.5D where a regular grid is populated with height data. This is also called a height map or relief map. This type of map only represents structure that is visible in a plan view. The mapping

data naturally takes this form when range measurements are made looking down from an altitude much greater than the relief of the scene. The result of this mapping pipeline is a full 3D point cloud which can be gridded any number of ways or displayed as a triangulated mesh. However, a 2.5D grid representation is a convenient framework for dividing up a point cloud for operations in this pipeline. It is also an intuitive way to view height information on a flat page so that is how the data will be presented in the results section.

3.3 Evaluation of Map Quality

Decisions on how to construct a map are informed by the map's ultimate application. For instance, producing maps for navigation requires conservative depth estimation biased towards the shoalest depth to comply with regulations [7]. The maps produced in this chapter are intended for quantitative scientific investigations of the sea floor.

A number of criteria related to the map application are important to the design of a mapping algorithm. The qualities that are considered in this design are as follows:

- Grid resolution. Higher point densities are important for resolving detail, so long as each point is contributing additional information. Greater point density allows a higher grid resolution if the points are accurately localized.

- Gaps. Gaps in the data make it difficult to interpret. If possible they should be filled with real data, even if it is at a lower resolution. Interpolation can also be used to fill the gaps, but interpolated data is generally of less value than real data. In an interpolated map, it can be difficult to distinguish between the two which can cause the user to be over confident.

- Artifact reduction. The user must be able to make precise measurements of individual features in the map. Artifacts such as distortion and 'ghosting', where obviously identical features are mapped in multiple nearby locations,

need to be avoided. Such misalignments are related to sensor calibration errors and navigation errors. A logical threshold for concern is when those errors are greater than the grid size allowable by the instrument's resolution.

- Preserving discontinuities and detail. Discontinuities in the terrain such as those related to man made artifacts or hydro thermal vent spires must be preserved. If sensors with two different sampling frequencies measure an area, its preferable to represent the scene with only data that has the higher sampling frequency. This avoids low pass filtering and a loss of information in areas of high relief area.

- Outliers. Both sensors produce outliers in the range data. A good mapping algorithm rejects these outliers without rejecting good data.

These criteria can be used to qualitatively asses the fidelity of a map. They address the more abstract side of map quality which directly contributes to how effective the map is for its specific application. The mapping algorithm described in this chapter was developed with these specific criteria in mind.

3.4 Methods

This section describes an algorithm to merge data from two sensors into a map of the sea floor. Specifically it focuses on combining range data from stereo cameras and multibeam sonar assuming known vehicle poses. The following are steps in the process which effect one or more of the above criteria. The way the steps are independently parametrized is used to maximize the map's quality according to the criteria.

3.4.1 Stereo

There are a number of ways to produce range data for mapping from two cameras. The previous chapter used sparse features to match and triangulate three dimensional feature points. This is a good approach for navigation refinement

because it reduces an image to its most unique features. There is no need to keep a record of the 3D position for every pixel of the image because only unique features are useful in data association.

A high point density is often desirable for mapping applications, and sparse feature matching cannot be used to determine a depth measurement for every pixel. A different approach to stereo matching called dense stereo correspondence is more capable of computing a depth for each pixel. Dense techniques are more suited to this task because the feature correspondence search is limited to a set of putative correspondences which lie on the epipolar line in the conjugate image. The way that matches are established using this constraint can vary greatly. A review and classification of current methods can be found in Scharstein and Szeliski [15].

The simplest of the dense methods is the Block Matching Algorithm implemented in OpenCV [16]. A window around a given pixel in the key image is compared using the sum of squared differences to likely pixel matches in the conjugate image. The correspondence search region is constrained by user input of the minimum and maximum pixel disparity range. Then correspondences are established within this range. Stereo correspondences found between pairs of images can be triangulated to for a 3D point cloud.

The Block Matching dense stereo algorithm is used here. It is fast and its various filters reliably reject outliers without a lot of tuning. Additionally, since no smoothing constraint is used, edges and textures are largely preserved. This consideration is important because the majority of dense methods make assumptions about the shape or smoothness of the environment which are appropriate for urban and indoor settings but are violated in natural terrain of the sea floor.

3.4.2 Multibeam

Multibeam sonar data requires somewhat less processing than stereo cameras to formulate 3D points. Much of the signal processing necessary for beamforming is done by the sonar's data acquisition software. After the data has been gathered,

the maximum intensity along a given beam is chosen as the distance to the sea floor along each of the 512 beams in a single ping. The ranges for each ping are naturally 3D point clouds which can be assembled into a map by projecting them into the map coordinate system via the navigation data.

During the processing we reject the outer 15 beams on each side as well as the center 10 beams. Both locations contain a large number of outliers. Rejecting a large number of beams is generally an aggressive an approach which rejects too much good data to be worth its simplicity. In this case however, setting an aggressive outlier rejection criteria is preferable for two reasons. First, since the survey geometry was designed to provide 200% overlap for the 45° Field of View (FOV) camera measurements, there is twice as much overlap for the 90° FOV multibeam sonar. This means that outlier rejection is unlikely to open up gaps in the map and at worst will simply cause a slight reduction in sampling frequency in areas where good data was incorrectly rejected. Second, this type of aggressive rejection makes the multibeam range data nearly free of outliers. This is very difficult to do for stereo range data making the multibeam data a good tool for rejecting bad stereo data. Ultimately, purging the multibeam data of outliers at the cost of losing some correct data appears to be worthwhile due to survey design and the difficult characteristics of stereo outliers.

3.4.3 Hybridization

Hybrid maps are created by selecting from the available visual or acoustic data to fill each grid cell of the map. The concept is that a map can be constructed by selecting the best data from a redundant data set by accounting for the specific characteristics of the sensors. The methods used here combine modalities with specific attention to the map qualities presented previously. The previous chapter focused on finding the sensor pose for each mapping measurement. This chapter follows by focusing on projecting them into a common frame using criteria to select the best data for each location of the map (Fig. 23). At this point, no additional

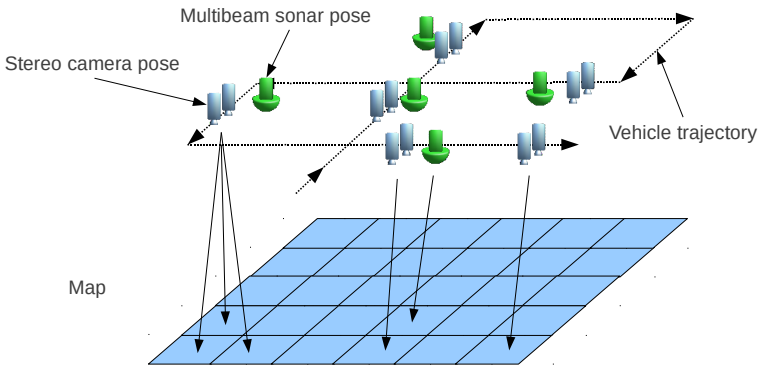


Figure 23. Mapping concept. Sensor poses are known at this point so the next step is to decide which data to use to populate each grid cell of the map. This chapter focuses on how to select the best data from each sensor with which to build the map.

navigation refinement will be done.

3.4.4 Simple Averaging

A simplistic method used for combining data from multiple sensors is to bin and average the data with no considerations made for outliers, misalignment or sensor characteristics. The area of the map is divided up into grid cells, 2×2cm. All points which fall into the cell are averaged to get the depth for that cell. This map is created for comparison (Fig. 24). It gives an idea of how the modalities might compliment each other, as well as demonstrating the specific problems which need to be addressed when combining their data.

3.4.5 Mapping based on local criteria

Averaging illustrates the dominant issues which arise from combining multiple modalities into a single map. Another approach is to cope with each of these issues individually, and select the best data for map assembly using criteria which are evaluated only over the grid cell in question or a small surrounding area. Initially, an appropriate grid size must be selected, then outliers and errors can be dealt with on a grid cell by grid cell basis. There are two main types of errors that can appear in a map. The first set of errors are large outliers from erroneous mapping sensor data, the second are more subtle errors related to sensor calibrations and

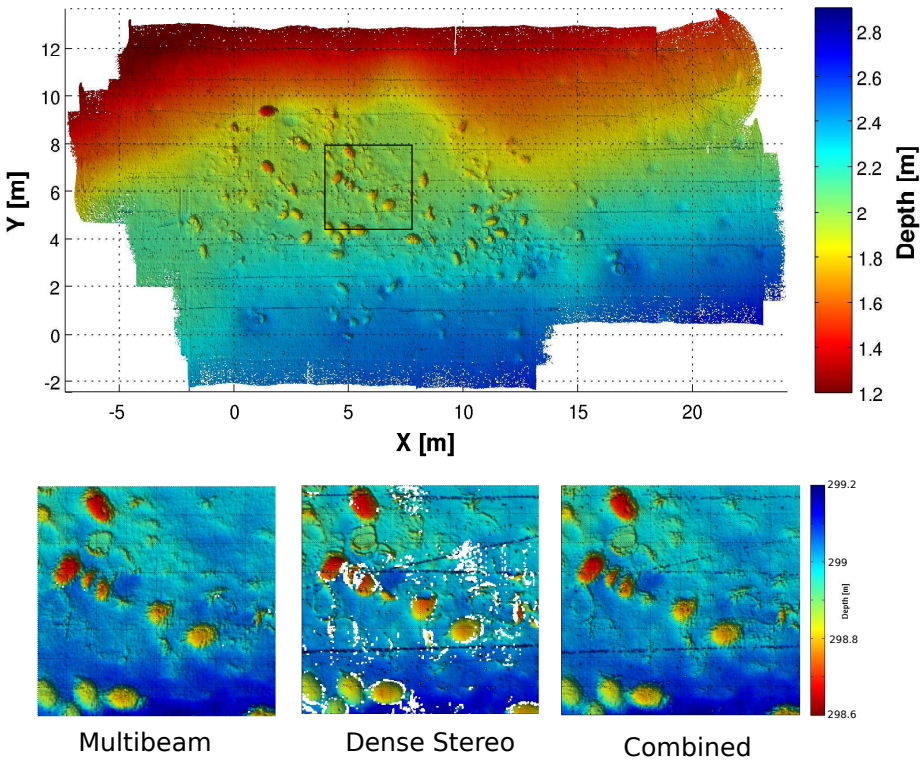


Figure 24. Averaging all sensors to construct multi-modal map. This approach makes no accommodation for outliers, different modalities, or misalignments. The hybrid map shows artifacts related to each of these issues. While we are able to fill in the holes usually seen in stereo, the precision of the stereo is degraded by averaging in the multibeam data. Additionally, the stereo outliers persist in the final map degrading the relatively outlier free multibeam range data.

remaining navigation errors.

Grid size

Grid size is selected to trade off between accuracy and resolution. There is no one grid size which is perfect for a given map. Instead trade offs must be considered as the map is being constructed. The minimum grid size is related to either the smallest grid size which still consistently contains one or two data points, or the smallest grid size which doesn't show obvious artifacts or errors. The smallest grid size that contains real information will have the highest spatial resolution. However, due to inherently noisy measurements, accuracy is improved when you can average over more measurements. This is achieved by having larger grid cells containing many points.

Properly trading off accuracy and resolution requires a bit of tuning. First we select a grid cell size that tends to contain result in as many grid cells occupied with range data as possible. We then grid the point cloud to the chosen size and compute the gridding confidence. An appropriate gridding confidence threshold is set and then the map is assembled. If too much of the map is cropped out, either decrease the gridding confidence (which will increase the likelihood of ghosting) or increase the grid cell size, depending on which is more valuable.

It is important to note that much of this discussion makes the assumption that most error in x and y is from navigation and most error in z is due to intrinsic sensor errors. Sensor calibration errors can also manifest in x and y but these simplifications are still a reasonable tool for selecting a grid size and pruning maps associated with poorly constrained vehicle poses.

Egregious outliers

Each sensor has outliers with particular characteristics. Taking these into account, its possible to reduce their effect on the final map. In particular, stereo outliers can be very difficult to eliminate without manually adjusting rejection

thresholds. On the other hand, multibeam ranges are relatively outlier free. This information can be leveraged to create an effective outlier rejection scheme which requires minimal input on the part of the user.

Dense stereo based outliers are usually the result of poorly matched pixels between the left and right images. These errors do not follow a normal distribution around the true range value. Instead they are often very far from the true range value. These commonly occur at the edges of the image and often there often many more of these outliers in a single grid cell than there are good measurements from either sensor. Areas of low texture which are frequently a problem for stereo matching are filtered out by the Block Matching stereo algorithm and generally don't result in outliers. The standard methods for rejecting stereo outliers are to remove points who's matches don't conform to the epipolar constraints of stereo system. However, dense stereo imposes the epipolar geometry as a constraint on matching, thus mismatches already conform and the constraint is not effective for outlier rejection.

The multibeam sonar has comparatively few outliers per grid cell is a useful tool for identifying stereo outliers. We compute the median and the square root of second moment about the median (standard deviation) of sonar range values in a four cell radius around a cell. The median is used instead of the mean to reduce the influence of any multibeam outliers present in the cell. Any stereo measurement more than three standard deviations from the median is rejected. There are many fewer multibeam outliers than stereo outliers and those that exist are rejected later in the process. As a result, it is not necessary to explicit reject them here. However, when they are present, the use of the median keeps them from having undue influence on stereo outlier rejection (Fig. 25).

Subtle errors

Once any egregious errors have been rejected, errors due to navigation, calibration and fundamental differences between sensors must be dealt with.

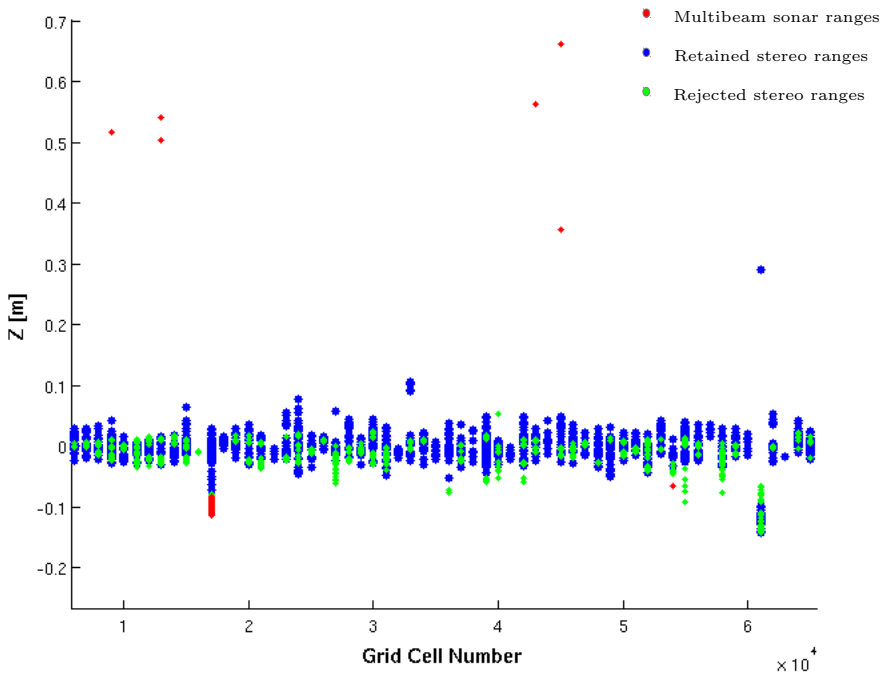


Figure 25. Stereo vision outlier rejection using multibeam sonar. Depth measurements normalized by sonar neighborhood median are shown. Stereo ranges are in green if kept and red if rejected. The sonar ranges are shown in blue. In the next step, any sonar that occupies the same cell as a stereo point will be rejected which eliminated the sonar outliers shown here.

- **Gridding Confidence Based Selection**

  Ghosting will result from misalignments between submaps. This is not explicitly addressed in the outlier rejection step. Much of ghosting is a result of errors in navigation which aren't fully resolved during navigation refinement. Certain areas of the map might contain fewer constraints on navigation making them not adequately correlated with the rest of the map. Therefore navigation refinement may fail to bring the maps into very close alignment.

  The amount of navigation error in x and y, and to a lesser extent z determines the appropriate grid size. With high resolution mapping sensors, navigation error is often the dominant error source, and you might expect to see ghosting if your grid size is smaller than navigation error. Instead of increasing grid size to accommodate navigation error at the expense of resolution, poorly aligned submaps are detected and removed, allowing sensor resolution to dictate the grid size at which a self consistent map can be achieved. Even so, the minimum grid cell still should not be smaller than the minimum point cloud density. This avoids holes in the final surface reconstruction.

  It is reasonable to assume that a pair of overlapping maps with poorly correlated poses will be poorly aligned with each other. Using this assumption, we can rid the map of submaps which aren't consistent with each other by flagging poses which are poorly correlated with each other. To do this, each point in the multi-modal point cloud is assigned to a grid cell. The marginal covariances are computed between each of the poses which have maps present in the grid cell. This covariance is used to compute a percent confidence that those two poses have been localized correctly to within the size of a grid cell. If the gridding confidence is lower than a threshold, both poses are flagged. After all the poses contributing to each grid cell have been flagged, the flags are summed. Incrementing through each grid cell again, the map associated with the pose which has been flagged the most times for poor pairwise con-

fidences is eliminated from the grid cell. This process reduces conflicts by eliminated maps which aren't well correlated with respect to each other and keeps poorly constrained points out of the map.

Computing the gridding confidence starts with determining the marginal covariance Σ_{ij} of the x, y components of the transform $\mathbf{x}_{i,j}$ between two poses $\mathbf{x}_i$ and $\mathbf{x}_j$. $\mathbf{x_{i,j}} = [x, y]^\mathsf{T}$ is a Gaussian random variable described by the ellipse

$$\left[\mathbf{x} - \mu_{x_{i,j}}\right]^\mathsf{T} \Sigma_{i,j}^{-1} \left[\mathbf{x} - \mu_{x_{i,j}}\right] = k^2. \tag{7}$$

k^2 is a χ_2^2 random variable which parametrizes the ellipse. Setting k^2, to a value corresponding to the required level probability (α) gives the equation of the ellipse defining the error circle for that level of confidence. For instance,$k^2 = 5.99$ has a 95% probability or $\alpha = .95$ according to the χ_2^2 probability distribution function. If the entire ellipse falls within a square the size of a grid cell, that indicates that the the two points come from maps are correlated enough that there is α confidence that they truly exist within the same square (Fig. 26).

- **Sensor Selection**

  Attempts to fuse the two data sources together can result in reduced precision. The edges of objects which are sharp in stereo reconstruction become blurred, and surfaces which are smooth become rough. To address this, only a single sensor is used to compute the depth at a given cell. In general stereo range data is preferred, and sonar data is rejected whenever it shares a cell with good stereo range data.

3.5 Results

This processing pipeline is designed to aggregate data from two sensors to produce a map which best addresses the map quality criteria introduced in Section 3.3. These criteria are more abstract than the quantitative error metrics used in

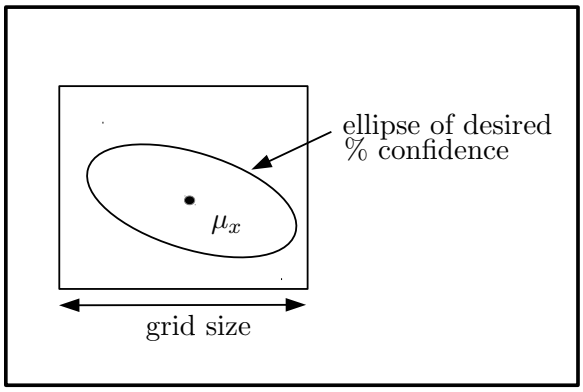


Figure 26. Confidence that maps appearing in the same grid cell are actually in same grid cell. This gridding confidence is computed using the marginal covariance of the two submap poses in X and Y. If the confidence of 95% confidence falls within a square the size and orientation of a grid cell, then those maps have an acceptable amount of relative uncertainty and both will be used in the grid cell. Otherwise, both will be flagged and the map associated with the pose related to the most bad flags will be rejected.

Chapter 2 but they are relevant because they are predicated on map characteristics important to the end user. Furthermore, these criteria can be used to evaluate the pipeline by comparing the criteria-based quality of each single sensor map to the final multi-modal map. This section begins by presenting the single modality maps and their characteristics. The parameters of the pipeline are explained in terms of their effect on the composite product. Finally the results of the single and multi-modal maps are compared.

3.5.1 Multibeam

Multibeam maps created directly from iSAM refined navigation have been investigated by both Kunz and Vaughn [13, 17]. The multibeam map in Figure 27 is the result of binning the point cloud of multibeam range data into grid cells and averaging the z values of the points in each grid cell. This map illustrates the type of artifacts which arise in multibeam maps and need to be addressed through multi-modal mapping. The artifact shown in the left inset commonly occurs when the vehicle stops briefly and many noisy multibeam pings are averaged together. In this case, the vehicle stopped and the navigation data dropped out for 6 seconds.

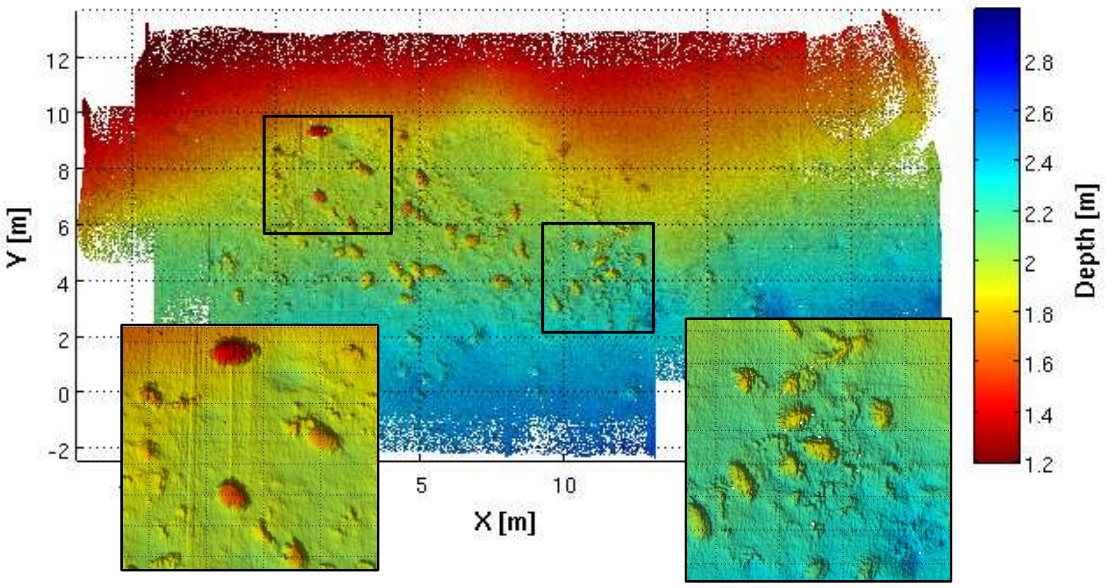


Figure 27. A multibeam map with 1.5cm grid size. The map appears reasonably self consistent with few gaps in the data. The point densities are much sparser in areas which have only been passed over once. So some grid cells around the edge are not populated, leaving holes. The left inset illustrates a linear artifact of the vehicle being stationary and a dropout of navigation data. The right inset shows the effect of noisy data and slight misalignment between overlapping submaps on the map.

This impacted the navigation data in a way this isn't modeled by the pose uncertainties during navigation filtering, therefore it is not properly dealt with during the navigation refinement. Because the vehicle pose covariances do not capture the uncertainty, it is difficult to identify and reject these bad points. In the right hand inset the edges of objects are not always distinct where multiple submaps overlap with slight misalignments. This results in blurry edges and repeated or 'ghosted' objects. Additionally, surfaces of amphorae which should appear smooth are often bumpy because of the noise in the range data.

Gaps in the data are another issue in this map. The effect of low point density can be seen around the edges of the map. Depending on the grid cell size, areas of the survey with no overlapping coverage may not have high enough point density to guarantee points will occupy every cell. This leads to sporadic empty cells. However, it may be advantageous to maintain this smaller grid size at the expense of small gaps in order to take advantage of the finer resolution available in regions with more overlapping coverage.

In spite of these issues, the multibeam map has some very favorable characteristics in terms of map quality. While there is some blurring of details and discontinuities, the data has few large misalignments. Small gaps are only evident where there is no overlapping coverage and this occurs mainly around the edges of the map and there are no large areas where the sensor fails to provide data. There are also few large outliers.

3.5.2 Stereo

A map assembled from dense stereo matching is shown in Figure 28. The most apparent feature is the number of gaps in the data. In this case, the gaps are caused by a poor calibration which prevents adequate alignment between the images during stereo matching. As a result, portions of the image could not be matched so range data could not be computed. Calibration issues are not the only cause of stereo ranging failures however. A number of other failures common to

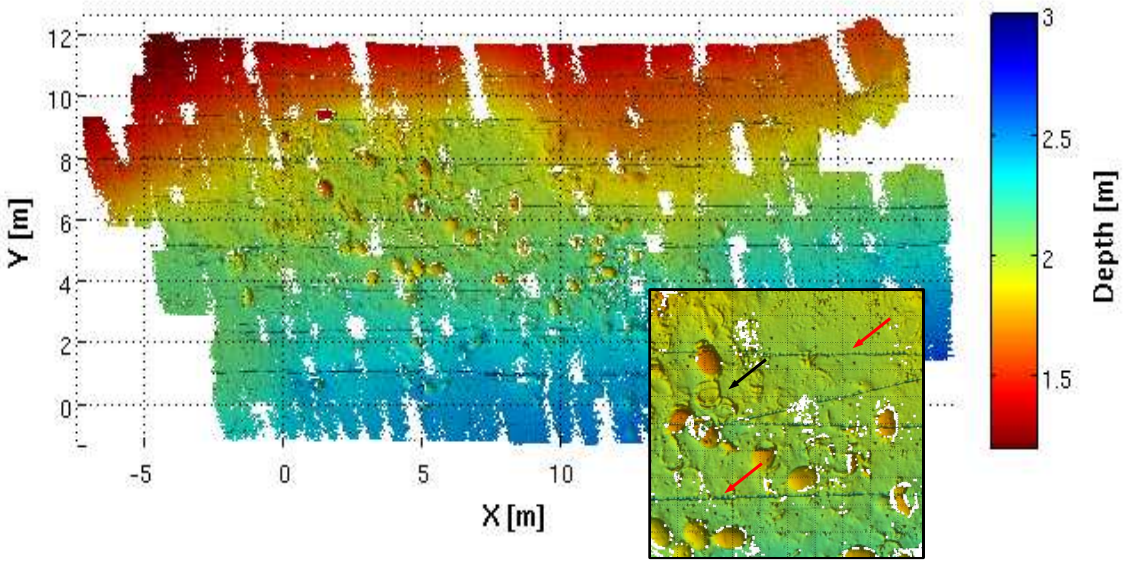
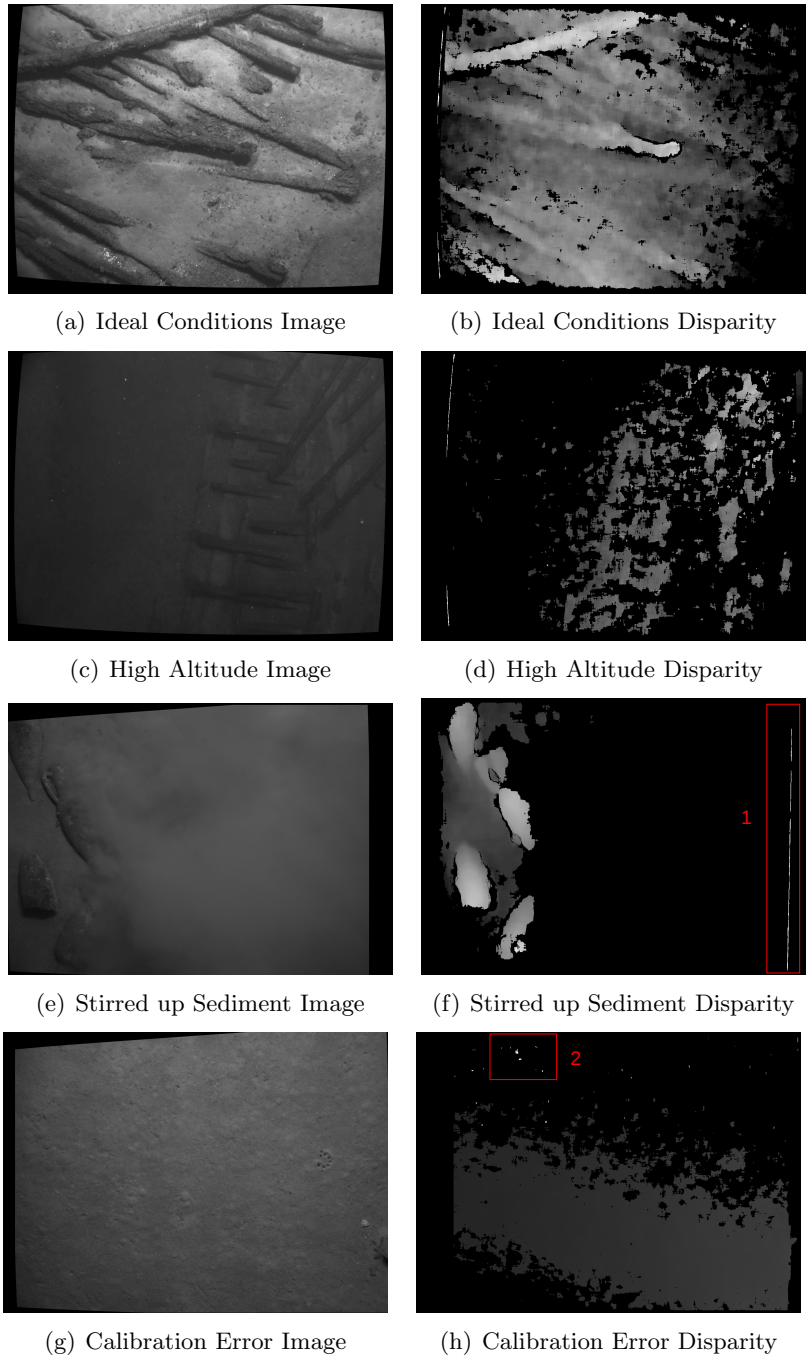


Figure 28. A stereo map with 1.5cm grid size and vertical averaging in the z. The precision of stereo measurements is apparent in the inset where the surfaces of amphorae are rendered smoothly. However, the gaps in the data are an obvious weakness. Outliers also appear in the inset (linear artifacts indicated by red arrows) and ghosting indicated by the black arrow.

underwater stereo are illustrated in Figure 29. These include high turbidity and high altitude, both of which complicate feature matching.

Figure 29(a) demonstrates that for ideal conditions the selected stereo matching algorithm, Block Matching, performs well enough to produce range data for mapping. It also generally decays gracefully as conditions decline. Few mismatches appear in poorly aligned or poorly textured regions leaving only gaps in the data. The exception to this is at the edges of the images where the distinct line between the image border and the black background can cause a large number of false matches (Figure 29(f), box 1). These matches can be difficult to reject without direct user intervention since they are not flagged by any standard automated stereo outlier rejection technique. Often these types of outliers can masked out, but instead, we opt to eliminate them during a later stereo outlier rejection step which simultaneously deals with outliers due to other types of false matches (Figure 29(h), box 2).



(a) Ideal Conditions Image (b) Ideal Conditions Disparity

(c) High Altitude Image (d) High Altitude Disparity

(e) Stirred up Sediment Image (f) Stirred up Sediment Disparity

(g) Calibration Error Image (h) Calibration Error Disparity

Figure 29. Dense stereo matching under varied conditions. This figure shows the left hand image of a rectified stereo pair (on the left) and a depth map produced using stereo matching (on the right). The Block Matching algorithm used here works well under idea conditions for underwater stereo. It fails however, for situations where altitude is high (or outside of calibrated zone of sensor), sediment has been stirred up, or the camera calibration is bad. Under the best conditions, the center of the image matches well but the corners generally do not.

Another notable feature of dense stereo matching is that has trouble resolving the edges of objects. This issue can be observed in the inset of Figure 28 where there is no data at the edges of many amphorae. There are two reasons for this. Some pixels are not matched simply because they are occluded, not visible in both the left and the right image as mentioned in Section 2.5.1. The other reason is that pixel matching is based on correlation between patches surrounding the pixels of interest. This assumes that the area around the pixel being matched is flat enough that parallax will not effect the content of the patch. This local flatness assumption is violated at the edges of objects. This is a common problem which applies to a greater or lesser extent to most stereo algorithms. Similarly, multibeam sonar has a limited ability to resolve edges due to having a relatively large beam footprint and as well as occlusions.

The stereo range data has much higher data density than the multibeam sonar, so it is better able to represent small details. However, the high point densities quickly become impractical for processing in Matlab. To deal with this, the dense stereo reconstructions were subsampled to a sample density slightly higher than the multibeam data density. Even so, the stereo cameras appear to provide better measurement fidelity. Figure 28 illustrates this where amphorae appear smoother and more distinct in the stereo map than in the multibeam map (Figure 27). Sharp sherds are also reconstructed faithfully in the stereo whereas they tend to be blurred out by the large beam width of the multibeam. The higher precision data available from stereo is good for representing detail but also makes ghosting more apparent. This is visible in in the inset in Figure 28 at the arrow.

3.5.3 Parameterizing the pipeline

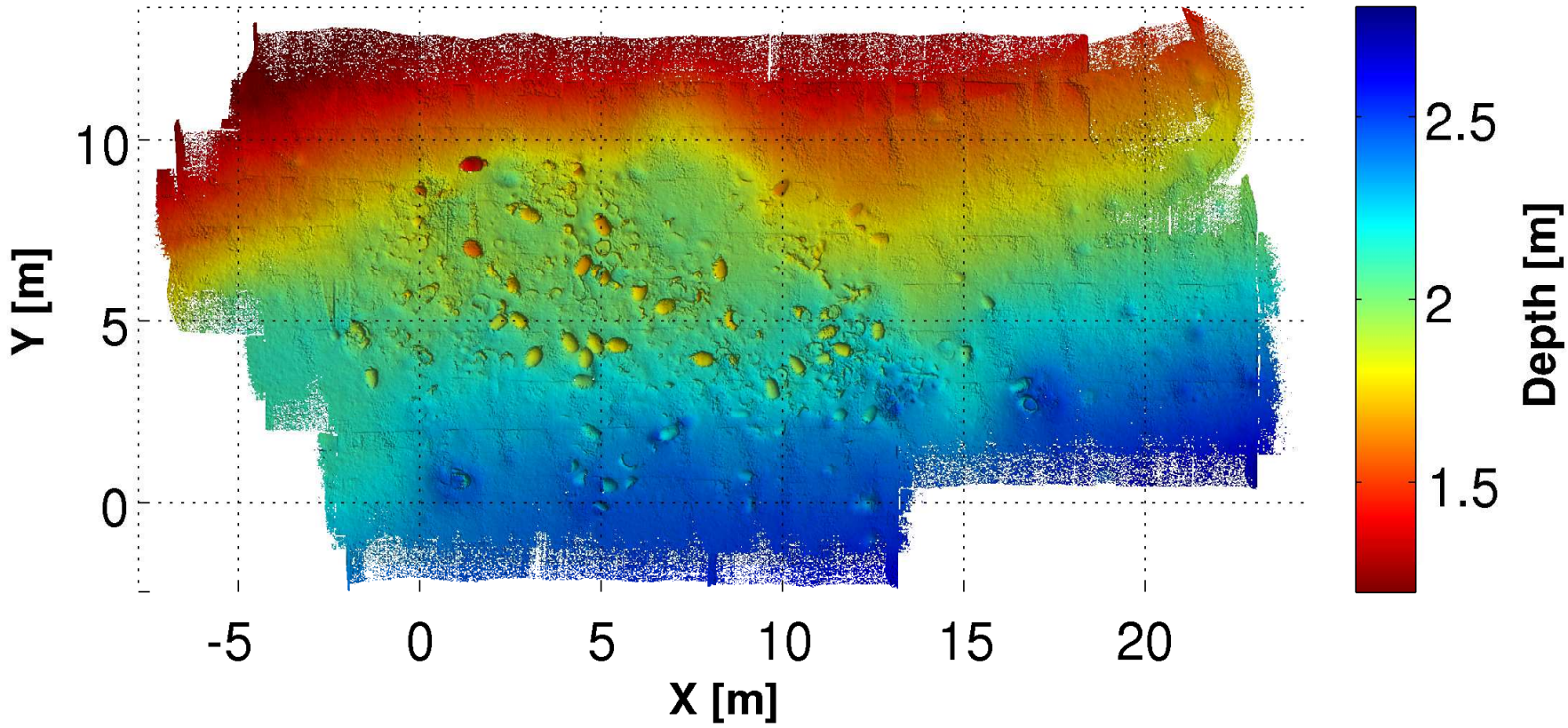


Figure 30. The final multi-modal map. This map is constructed from both stereo and multibeam sonar ranges, combined according to the procedure outlined in Section 3.4.5

The final map is shown in Figure 30. This map represents the results of the steps outlined in the in Section 3.4.5. The following sections summarize the results of the various steps, their parametrization and its effect on map quality. First a reasonable grid size is selected and validated, then the point cloud is gridded. Then several variables which parametrize outlier rejection are tuned and applied to the gridded point cloud. The outcome of this process compares favorably with the single modality maps.

While this map has no issues with ghosting, because there are plenty of links between poses, ghosting is a consistent issue in creating sea floor maps and it can occur any time there are not enough link constraints between poses with overlapping map data. To illustrate this point and the way that the mapping pipeline addresses it, the links between the crossing line and the rest of the survey were removed. This makes the crossing line poorly constrained and results in some ghosting where the crossing line overlaps the rest of the map.

Determining Grid Resolution

Grid size is the first parameter which must be set because then the point cloud can be gridded and all subsequent steps can be executed cellwise. This step helps decide on a reasonable minimum resolution at which to operate, however, the final result is a point cloud which can be re-gridded using any algorithm at any resolution.

Deciding on an appropriate grid resolution begins with choosing a minimum grid size that generally guarantees each grid cell will contain data. By plotting the percentage of occupied grid cells over a number of grid sizes, the minimum grid size becomes apparent. This is the point where increasing the grid cell size no longer increases the percentage of occupied grid cells. Figure 31 shows the percent occupancy of the grid for each sensor. It is necessary to examine both sensors. The sensor indicating the largest grid size should dictate the overall minimum grid size. By examining this plot, a reasonable minimum grid size of 1.5cm can be selected.

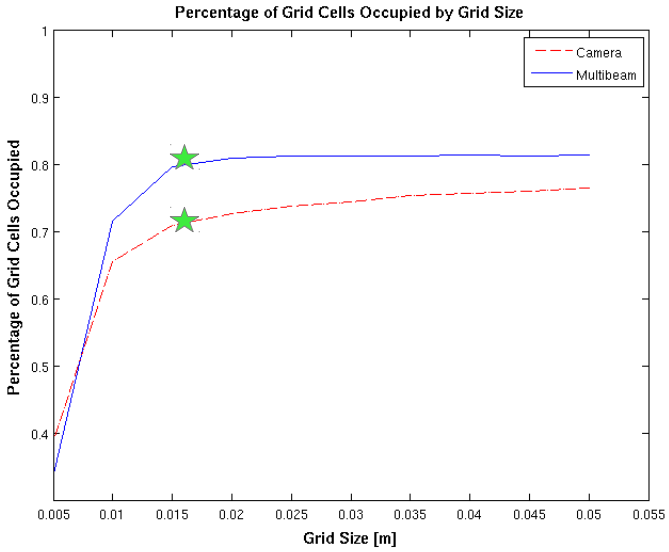
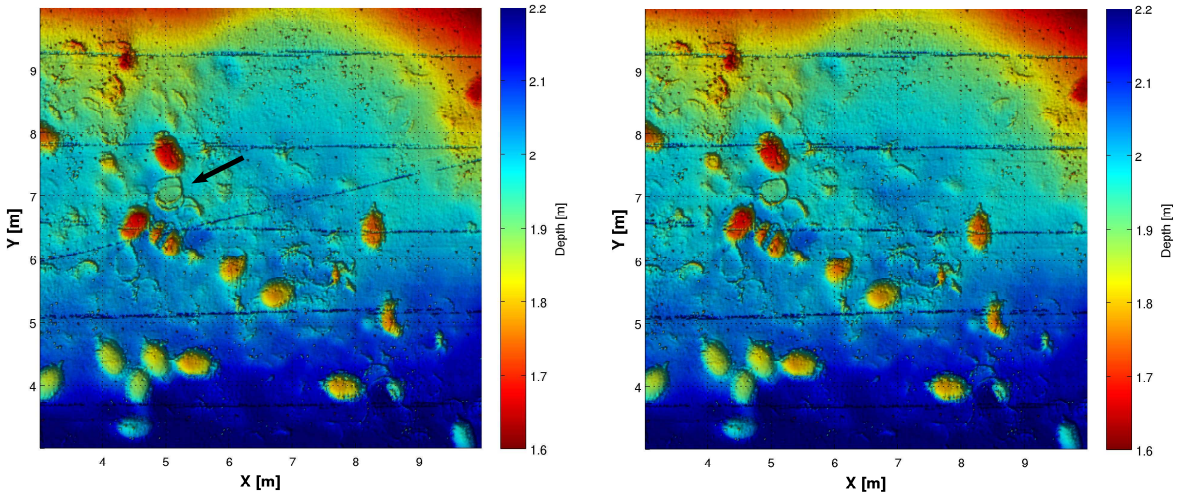


Figure 31. Using number of occupied grid cells for minimum grid size selection. This figure shows the percentage of occupied grid cells as a function of grid size. As grid size increases, the gain in grid occupancy levels off. The point where it levels off is a reasonable minimum grid size and is indicated by the star.



(a) 1.5cm Gridding

(b) With gridding confidence rejection

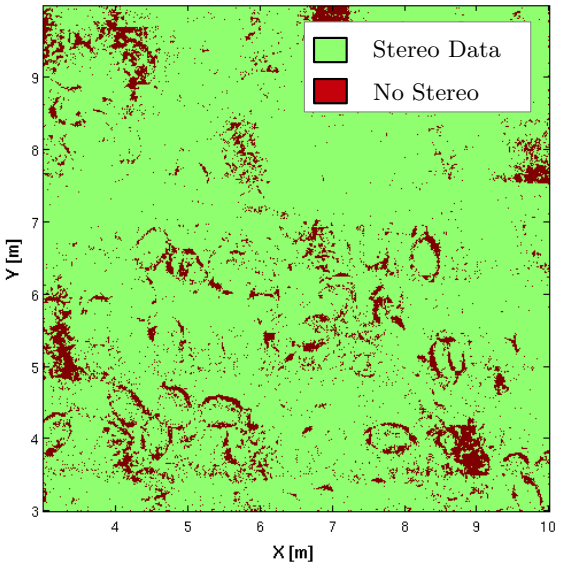
Figure 32. Verifying grid size selection on an averaged multi-modal map. **(a)** shows that some navigation error is still visible in the map at 1.5cm grid size. By using gridding confidence elimination and finding that it effectively removes the ghosting **(b)**, and re-averaging, we can demonstrate that the ghosting was due to a few poorly constrained poses instead of a pervasive high level of navigation error across the map.

Grid size trades off several map attributes. A larger grid size results in lower resolution but fewer apparent artifacts and fewer gaps in the data. Our current choice of grid size is strictly based on gaps. It also must be validated in terms of navigation error artifacts. If navigation error is widespread and obvious at the current grid size, then the grid size should be increased. If there only a few localized errors, these might be due to isolated poor navigation data and can be eliminated in an outlier rejection step without having to increase grid size.

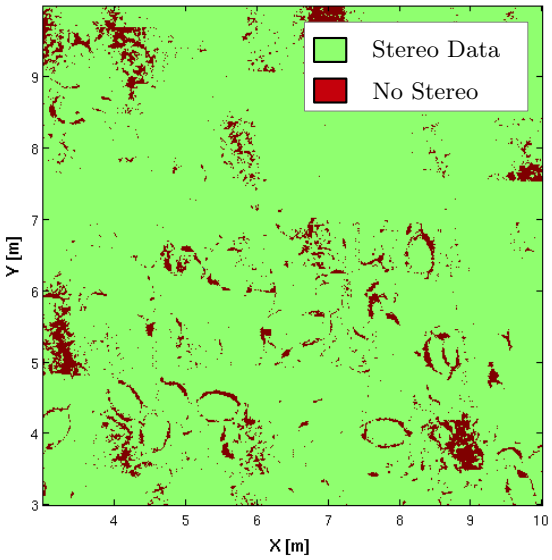
Errors related to navigation data can be observed by gridding and averaging in z (Fig. 32(a)). When this point cloud is gridded, there is clearly some ghosting in the area of the map indicated by the arrow. The question is whether this error is related to a small area of poorly constrained poses which can be eliminated as outliers, or if it is the dominant error magnitude for the whole map. This can be determined by eliminating points from different maps which don't have 95% confidence of being in the same grid square. In this case the gridding confidence elimination rejects the misaligned maps without reducing overall map quality (Fig. 32(b)). The resulting increase in clarity confirms this grid size is reasonable. If the resulting map had still contained misalignments or become noisier due to gridding confidence rejection, it would have been likely that the navigation error was too large for the chosen grid size. In the latter case, the grid size must be increased until it is on the order of the navigation error. There is no further trade off associated with a larger grid size other than the loss of resolution due to subsampling.

Outlier rejection

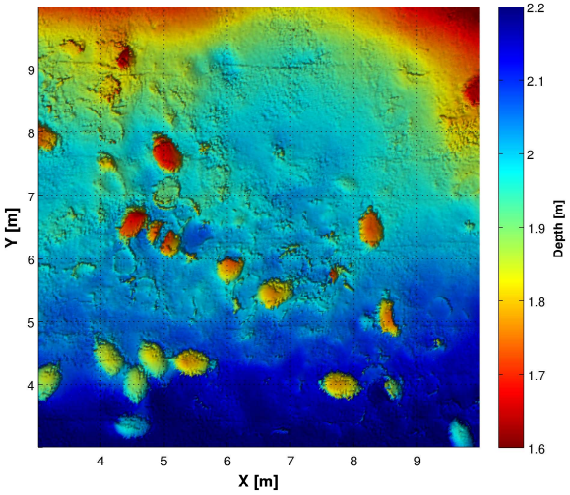
Outlier rejection is the next step in mapping once the appropriate grid size has been determined and verified. The main purpose of this step is to identify and reject egregious outliers. Misalignments and more subtle errors will be addressed during subsequent steps. Having observed that there are very few egregious outliers in the multibeam data, its reasonable to use this data to reject stereo data which contains far more large outliers.



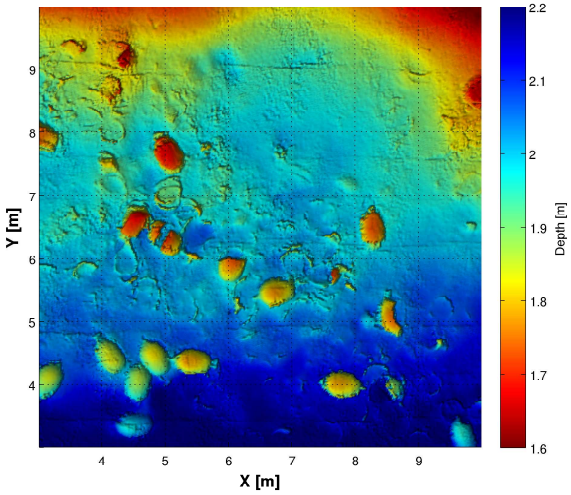
(a) Stereo occupancy after outlier rejection radius 0



(b) Stereo occupancy after outlier rejection radius 2



(c) Map after outlier rejection radius 0



(d) Map after outlier rejection radius 2

Figure 33. Demonstration of rejecting stereo outliers using multibeam cellwise statistics. The results of computing the rejection statistics from a single cell are shown in plot **(a)** and from a two cell radius in plot **(b)**. There is more widespread camera rejection when only one grid cell is used, and this corresponds to a bumpier looking map as observed in **(c)** relative to **(d)**. No other types of outlier rejection were used at this stage, these plots are strictly the result of rejecting camera data based on cellwise statistics.

Initially, outlier rejection was done simply by eliminating any stereo points lying more than three standard deviations from the median of the multibeam range for a given cell. This resulted in some valid stereo data being rejected simply because the multibeam data was noisier than the stereo or slightly misaligned. This appears as clipped off data creating jagged edges on objects, or widespread elimination of stereo data. This can be corrected by computing the rejection statistics using a two grid cell radius around the cell where rejection is being performed.

When using a two cell radius there is no evidence of rejecting valid data which would necessitate a larger radius. Meanwhile, the stereo outliers appearing as horizontal lines are consistently rejected (Figure 28, red arrows). This radius can be tuned for each new map. When alignment between the two modalities is nearly perfect, the rejection threshold can be computed from a one grid cell radius. Computing statistics from zero radius (or a single grid cell) isn't advisable since a single grid cell runs the risk of containing very few points and can produce statistics which poorly represent the area due to the noisiness of multibeam data. Figure 33(a) shows that when the rejection statistics are computed from only one grid cell, many sporadic camera points are rejected. These same isolated areas are not rejected when a two grid cell radius is used in Figure 33(b) resulting in smoother object surfaces in Figure 33(d).

Gridding Confidence Rejection

After rejecting large outliers, the more subtle misalignment errors can be addressed. Gridding confidence rejection can be used to eliminate points related to poorly constrained navigation data from the outlier rejected point cloud. While this step was run previously as a way to verify the grid size selection, now it is used as part of the map making pipeline.

Reducing ghosting can be accomplished by removing maps which are projected from poorly localized poses. In the example map, the crossing line is poorly con-

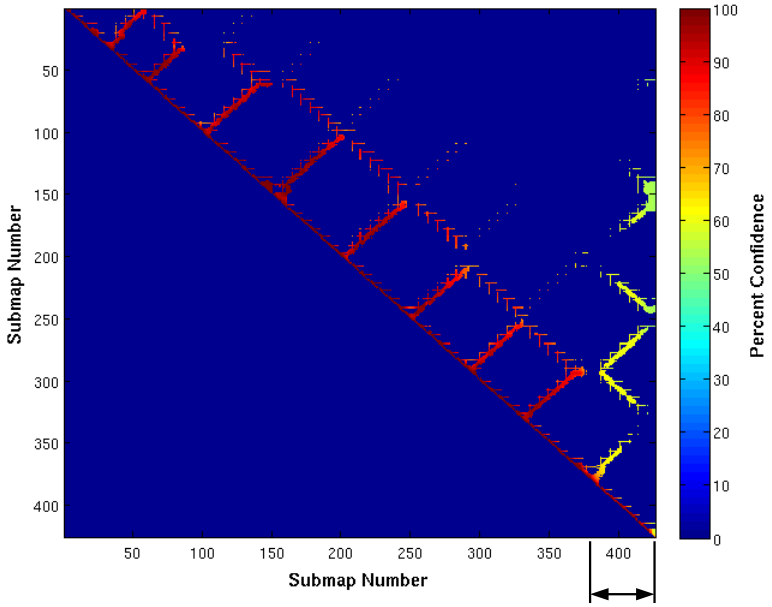
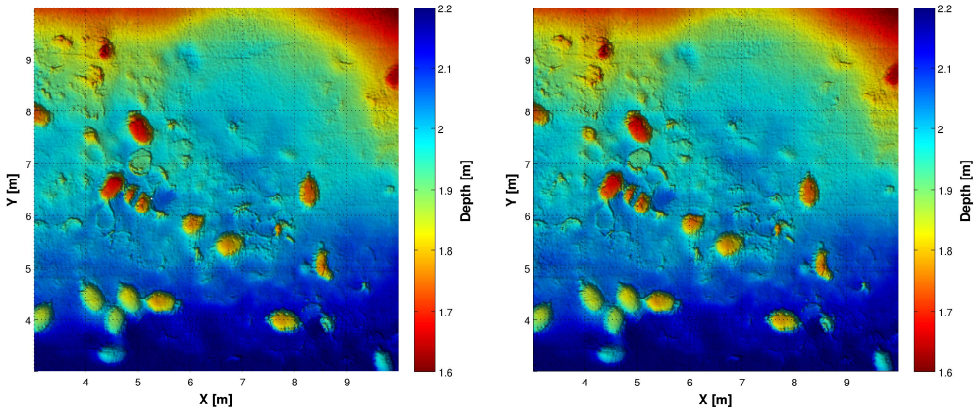


Figure 34. Gridding Confidence for overlapping submaps. This figure shows the percent confidence that poses contributing maps to the same grid cell are localized with less than a grid cell of uncertainty. Poses on the crossing line have low confidence of contributing points to the proper grid cell. Points projected from the poorly localized poses are good candidates for rejection since they may create ghosting artifacts due to misalignment.



(a) Outlier Rejected (b) Outlier & Gridding Confidence Rejected

Figure 35. Results of gridding confidence rejection. Removing the points related to poorly constrained navigation from the map removes ghosting and improves the clarity of the map.

strained to the map because the crossing line has no links with the main part of the survey. This lack of links leads to a relatively large marginal covariance between poses on the crossing line and poses from nearby portions of the map. In fact, the marginal covariance between poses on the crossing line and poses with overlapping submaps leads to only a 65% or lower confidence that points measured during the crossing line are being projected to the correct grid cell (Fig. 34). These gridding confidences are a good predictor of ghosting. The 95% confidence interval is both reasonable in theory and gives good results in practice (Fig. 35).

Generally, eliminating points from the grid cell average would tend to reduce accuracy because there are now fewer noisy measurements to average over. However, by eliminating only those points which are poorly localized, the map will be improved since points which are not representative of the surface in that grid cell have been removed. The exception to this is if only one point remains in a grid cell after rejection and that one point poorly localized. In this case, artifacts may continue to appear. At this point, the map the grid size should be increased, however this issue would have been apparent during grid size validation, and addressed with a larger grid size at that time.

Sensor selection

A single sensor is assigned to each grid cell during the sensor selection step. This step occurs after all other outliers have been rejected. By saving this step for last, any gaps in the an individual sensor's data created during outlier rejection can be filled with the remaining sensor. This step assumes that all the data which remains in a grid cell is accurate but that having only one sensor per grid cell is preferable. The stereo camera range data is higher resolution and more precise so it is selected whenever two sensors occupy the same grid cell.

There is some error apparent between the two modalities which manifests itself at transitions between multibeam and stereo coverage. It appears to be a bias in z of $1 - 2$cm with the multibeam ranges being slightly longer. The

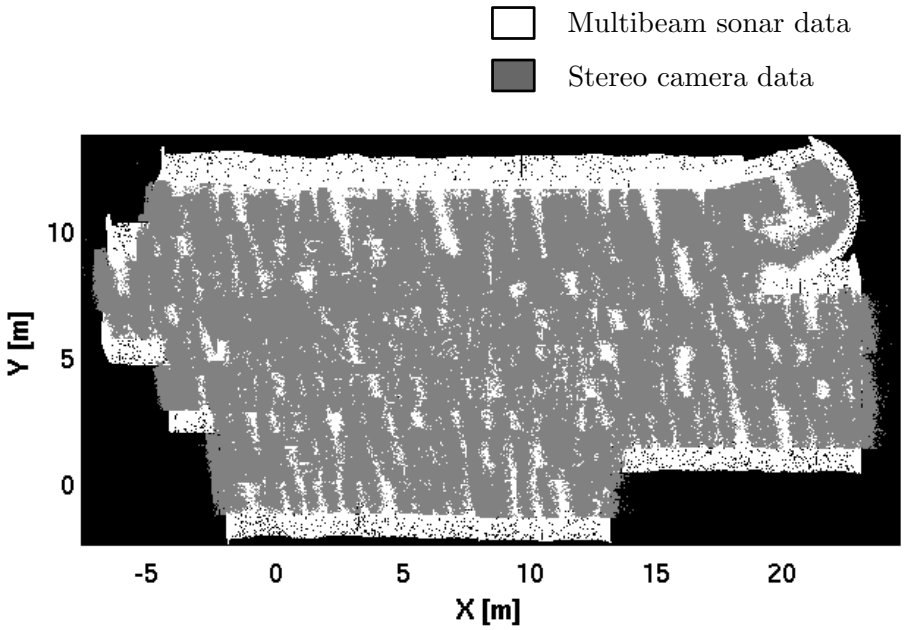


Figure 36. Sensor selection. Each grid square only contains a single sensor's data. This reduces the effect of averaging improperly aligned submaps and blurring the reconstruction.

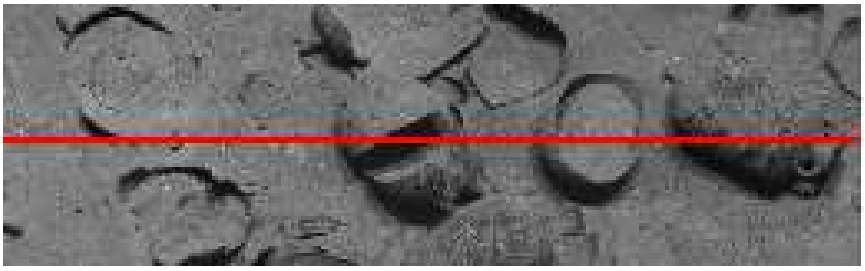
appearance of the bias is accentuated by the transition from the relatively smooth stereo measurements to the noisier multibeam measurements. A broader issue with this mapping pipeline is that there isn't perfect alignment between the camera and the multibeam data. Where there are many adjacent cells containing different sensors, the map appears bumpy even if the terrain is smooth (Fig. 30).

The sensor selection step allows the map to retain the favorable characteristics of the stereo data. The stereo data remains undistorted by multibeam data because the latter is only used in places where there is no stereo data (Fig. 36). Textures are also preserved which is not possible when using multibeam only or multibeam and camera data averaged together.

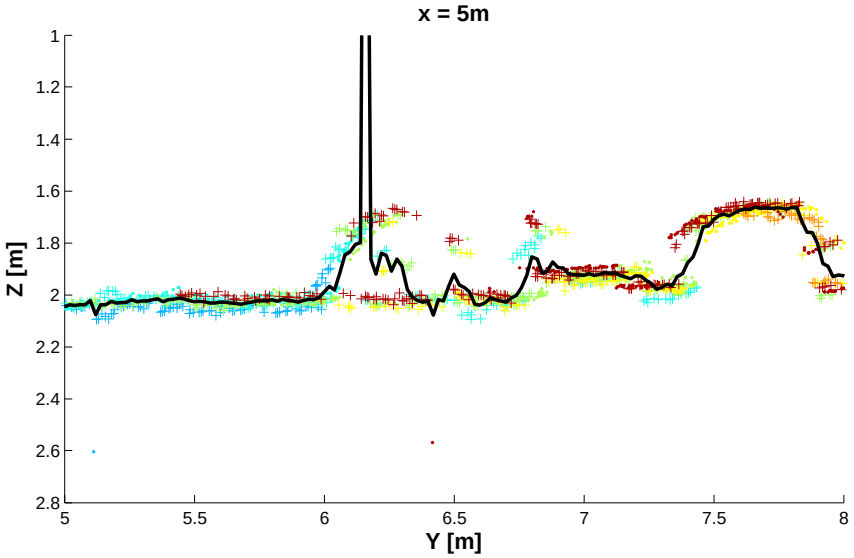
3.5.4 Comparison between single and multiple modalities

Comparing the single and multi-modality maps serves as a measure of performance for the mapping pipeline. Grid resolution was maintained at 1.5cm for the final map product. At this grid size, the multi-modal map is generally better than either single modality map.

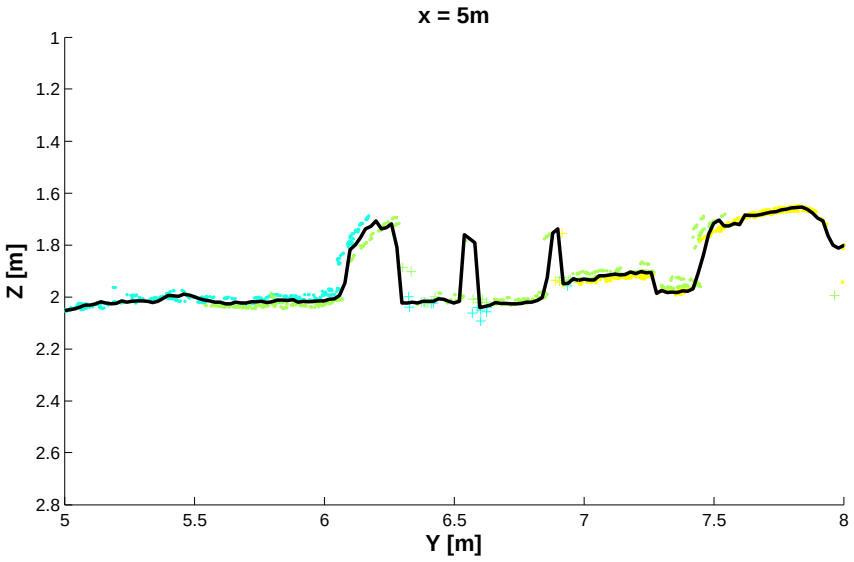
The multi-modal map effectively incorporates the strengths of the single modality maps. The gaps in the stereo data are corrected using the multibeam data. However stereo data is the dominant data source which results in a highly detailed map. Outliers have been successfully rejected with minimal user intervention. The remaining artifacts in the final map are largely those which were also in the single modality maps. The area mentioned in Section 3.5.1 for having some dropped navigation data still contains artifacts, there is also a remaining artifact due to roll bias in the stereo cameras. It is present in both the camera only and the multi-modal map and is a flaw in the navigation refinement as opposed to the mapping.



(a) Photomosaic



(b) Initial point cloud & estimated surface



(c) Final point cloud & estimated surface

Figure 37. Initial and final point cloud vertical slices. **(a)** Photomosaic of the area around a vertical point cloud slice. **(b)** The slice of the initial point cloud shows that it is several cm thick. This thickness can obscure small features. **(c)** The final slice shows the features more accurately and is more faithful to textures. The profile is smooth where the image is smooth and rough where the image appears rough.

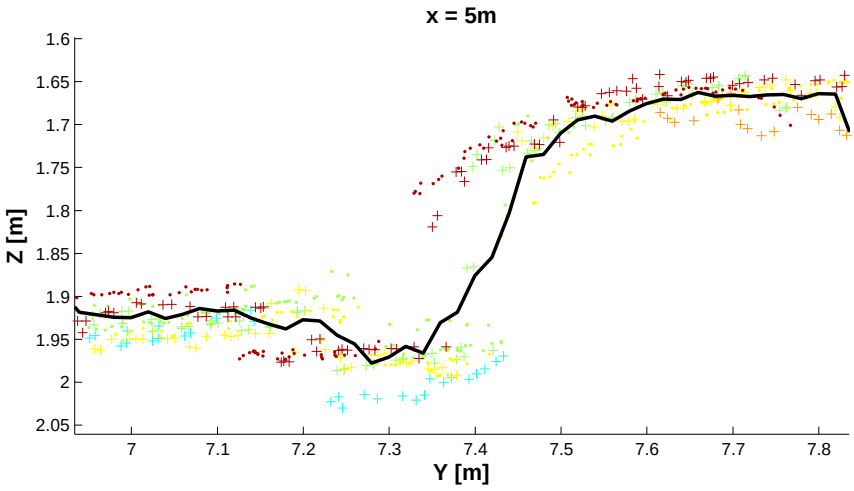


Figure 38. Profile of an unprocessed point cloud. The point cloud here shows both multibeam (+) and camera points (.). The colors correspond to the map number with warm colors being submaps from the beginning of the survey and cool colors from the end.

3.5.5 Point cloud profile

Figure 37(b) shows a slice of the naively merged multi-modal point cloud. The black line shows a profile of the surface created by gridding and vertically averaging. The initial point cloud is several centimeters thick and appears to have a multi-modal distribution in z, particularly in areas where there is more structure. A close up shows that even in flatter regions, the vertical distribution of points in a grid cell tends to be composed of clusters (Figure 38). Here submaps which are adjacent in time are also adjacent in the colormap, so warm submaps were acquired at the beginning of the survey and cool colors at the end. Notice that maps with similar colors tend to lie closer to each other than they do to other submaps in the same x, y location. This occurs regardless of modality. This indicates that navigation error is still a significant source of error in the map.

As mentioned previously, vertical averaging across such a clustered distribution in z will not give a good scene representation. The appearance of bumpiness in the initial estimated surface is mainly due to averaging across the unevenly distributed ranges in each grid cell. The final point cloud is much thinner, and it is able

to faithfully render smoothness where the images show the bottom is probably smooth.

3.6 Discussion

This pipeline produces maps synthesized from multibeam and camera data. The design of the pipeline was intended to address the map fidelity criteria laid out in Section 3.3. The steps above specifically address the issues of grid resolution, outliers, artifacts, detail preservation and gaps. This results in an improved map relative to the single modality maps. However, a few issues remain unresolved by this process. This section discusses the successful aspects of the pipeline as well as its limitations and potential remedies.

Gaps

There appear to be two types of gaps in the gridded data. One consists of large missing segments of data. These can be due to sensor malfunction or data elimination during outlier rejection steps. The solution to this problem is to substitute in other data. This pipeline is effective for coping with these types of gaps. The holes in Figure 28 are filled using multibeam data as shown in Figure 36.

The other type are small dispersed gaps. These occur when the grid size is too small for the sensor's measurement density leaving grid cells that are not populated with range data. If small holes are pervasive, it means that measurement density is too low for the chosen grid size and it should be increased. However, if increasing the grid size is not desired, or only small sections of the map have this problem, there are two ways these gaps could be addressed. First you could try to extract more range data from existing sources. Lowering filter threshold on dense stereo matching will result in more stereo matches, however these tend to be mainly bad matches. Therefore, this option impractical. Another option is to interpolate existing data. Interpolation should be handled with caution since it

creates data where there was none by making assumptions about the characteristics of the surface. This is inadvisable because interpolated data can be confused with real data and make the user over confident. However, when the gaps are the size of a grid cell, a local interpolation technique such as linear interpolation would help make the map more readable without running the risk of creating fictitious structures or flat ground where there is texture.

Accuracy

Evaluating the map accuracy is very difficult with sea floor data. To evaluate the proposed method for true accuracy would require a ground truth data set. Lacking this data set, we can note that the method requires no assumptions such a scene flatness, frequently used in photomosaicking, which introduces systemic distortions. Additionally, the map-to-map error (Fig. 22) shows that the map is relatively self consistent which also increases our confidence in the map's accuracy.

Artifacts

Many of the artifacts initially present in the naively combined multi-modal map have been eliminated, however a few biases remain. The roll bias in the camera data is still apparent after mapping, which can be expected since no specific effort was made to eliminate it in the mapping processes. This is clearly a shortcoming in the previous navigation refinement step. However, it may be possible to reduce the effect of a roll bias during mapping without adversely effecting the map. Roll bias is most apparent at the edges of submaps. In areas where multiple maps overlap, points which are farthest from the center of the submap can be eliminated. If a sophisticated gridding process is used to process the ultimate point cloud, blending techniques could be used to reduce the appearance of such artifacts [11].

There is a small difference in z value between the camera and the multibeam data. This is apparent in areas where there is no camera data and the multibeam data fills in the gap. In those places, there is a small depth discontinuity. While

its possible to reduce the appearance of problems like this with blending, that is strictly a cosmetic solution. This issue requires better agreement between the two sensors to be established during navigation refinement.

Sequential steps versus a unified approach to mapping

The presented approach is a series of steps that deal with individual mapping issues specifically and directly as they arise. A series of individual steps has the advantage that it is easy to add and subtract steps and tailor the algorithm to data sets with unique issues. More unified techniques could be harder to adapt in cases where they fail. Another advantage is that tuning the algorithm is fairly straightforward since available dials correspond directly with physical parameters. However, this design can be an issue because the approach doesn't solve problems that haven't been explicitly modeled. Instead specific techniques need to be developed to deal with the characteristics of some data sets. For instance, the outlier rejection technique used here rejects bad points based on their distance from the mean value of the grid cell. Over terrain with very large discontinuities, this method may not be able to distinguish outliers from large discontinuities so a new outlier rejection technique might be necessary. Even so, this method has addressed some common issues of mapping which can be expected to reoccur thus these techniques will generally transfer to other mapping problems.

The alternative would be to develop a more unified global method for surface reconstruction. Such a method was devised for this data set where multiple depth hypothesis were generated based on clustering algorithm for each grid square. Depth hypothesis were resolved using a Markov Random Field to minimize a cost metric which weights hypotheses based on factors influencing map quality. The challenging aspect is that tuning the cost function is very difficult. The tunable parameters tend to be highly abstracted from their physical effect on the map. Additionally, the smoothness constraint implicit in a Markov Random Field made over-smoothing a pervasive issue. The steps in the processing pipeline presented

here appear better suited because the tunable parameters aren't relative weights, they are parameters with physical interpretations.

This result shouldn't eliminate the use of tools such as Gaussian Processs (GPs) or B-Splines. These unified methods could be very powerful for representing the scene. GPs naturally lend themselves to adaptive grid sizes to reflect actual resolution of the available sensors and have been used successfully for multi-sensor mapping on land [18]. One issue with such approaches is that they require a very strong understanding of the error characteristics of each point in the map which are derived from good modeling of the sensors and navigation data. Having a limited understanding of interplay of these uncertainties will lead to results such as over valuing one sensor's data, failure to recognize noise or biases, over-smoothing or over-fitting. These approaches also present a significant computational burden but this will be less of a problem as methods and hardware improve. In spite of these issues, such unified methods are likely to be the next evolution in multi-modal surface reconstruction, now that we have a more thorough understanding of how the two sensors interact in a single map. Even so, the outlier rejection steps listed will still be needed in a unified approach to surface reconstruction.

List of References

[1] H. Singh, J. Howland, D. Yoerger, and L. Whitcomb, "Quantitative photomosaicking of underwater imagery," in *OCEANS '98 Conference Proceedings*, vol. 1, 1998, pp. 263–266 vol.1.

[2] O. Pizarro and H. Singh, "Toward large-area mosaicing for underwater scientific applications," *IEEE Journal of Oceanic Engineering*, vol. 28, no. 4, pp. 651–672, October 2003.

[3] R. Ballard, L. Stager, D. Master, D. Yoerger, D. Mindell, L. Whitcomb, H. Singh, and D. Piechota, "Iron age shipwrecks in deep water off Ashkelon, Israel," *American Journal of Archeology*, vol. 106, no. 2, April 2002.

[4] O. Pizarro, R. Eustice, and H. Singh, "Large area 3-d reconstructions from underwater optical surveys," *Oceanic Engineering, IEEE Journal of*, vol. 34, no. 2, pp. 150–169, 2009.

[5] T. Nicosevici, N. Gracias, S. Negahdaripour, and R. Garcia, "Efficient three-dimensional scene modeling and mosaicing," *Journal of Field Robotics*, vol. 26, no. 10, pp. 759–788, 2009. [Online]. Available: http://dx.doi.org/10.1002/rob.20305

[6] S. Barkby, S. B. Williams, O. Pizarro, and M. V. Jakuba, "A featureless approach to efficient bathymetric slam using distributed particle mapping," *Journal of Field Robotics*, vol. 28, no. 1, pp. 19–39, 2011.

[7] B. R. Calder and L. A. Mayer, "Automatic processing of high-rate, high-density multibeam echosounder data," *Geochemistry, Geophysics, Geosystems*, vol. 4, no. 6, pp. n/a–n/a, 2003. [Online]. Available: http://dx.doi.org/10.1029/2002GC000486

[8] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. The MIT Press, September 2005.

[9] M. Johnson-Roberson, O. Pizarro, S. Williams, and I. Mahon, "Generation and visualization of large scale 3D reconstructions from underwater robotic surveys," *Journal of Field Robotics*, 2009 (in press).

[10] B. Curless and M. Levoy, "A volumetric method for building complex models from range images," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM, 1996, pp. 303–312.

[11] P. J. Burt and E. H. Adelson, "A multiresolution spline with application to image mosaics," *ACM Trans. Graph.*, vol. 2, no. 4, pp. 217–236, Oct. 1983. [Online]. Available: http://doi.acm.org/10.1145/245.247

[12] H. Singh, C. Roman, L. Whitcomb, and D. Yoerger, "Advances in fusion of high resolution underwater optical and acoustic data," in *Underwater Technology, 2000. UT 00. Proceedings of the 2000 International Symposium on*, 2000, pp. 206–211.

[13] C. Kunz, "Autonomous underwater vehicle navigation and mapping in dynamic, unstructured environments," Ph.D. dissertation, MIT-WHOI Joint Program, November 2011.

[14] 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.

[15] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, pp. 7–42, April 2002. [Online]. Available: http://portal.acm.org/citation.cfm?id=598429.598475

[16] G. Bradski and A. Kaehler, *Learning OpenCV: Computer Vision with the OpenCV Library*, 1st ed. O'Reilly Media Inc., 2008.

[17] I. Vaughn, "Microbathymetry using self-contained navigation and simultaneous localization and mapping," Master's thesis, University of Rhode Island, 2012.

[18] S. Vasudevan, F. Ramos, E. Nettleton, and H. Durrant-Whyte, "Non-stationary dependent gaussian processes for data fusion in large-scale terrain modeling," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, 2011, pp. 1875–1882.

CHAPTER 4

Conclusion

4.1 Introduction

This thesis presents a method for producing sea floor maps from multiple modalities. This is motivated by the recognition that the two sensors commonly available on underwater mapping platforms have complementary strengths. With this in mind we present two part system. This system has identified some of the challenging areas of the multi-modal mapping problem and addressed them by breaking the problem into navigation and mapping components. First mapping measurements from both sensors are localized in a common reference frame while enforcing consistency between their maps using navigation refinement. Considering the relative strengths and weaknesses of each sensor, data is then drawn selectively from the two sensors to populate a map. This map exceeds either individual modality's map on a number of map quality criteria while coping with outliers and remaining navigation error.

4.2 Summary of contributions

The system as a whole was successful at producing a multi-modal map. Several developments were necessary in accomplishing this:

- **Navigation framework with cross modality registration.** A navigation framework was developed using the iSAM smoother. This framework built on the multi-modal navigation refinement system developed by Kunz by incorporating cross modality links between stereo reconstructions and multibeam submaps. The cross modality links emphasized consistency between the stereo camera and multibeam sonar maps.

- **Summary of relevant error metrics.** Several methods were used to evaluate the quality of the navigation refinement. These metrics included an

alteration to Roman's map-to-map error metric for quantifying the degree to which overlapping maps agree with one another.

- **Multi-modal map fidelity criteria.** Evaluating a map's utility requires some more abstract criteria than error metrics presented in the navigation chapter. To evaluate the fidelity of the maps produced here, several characteristics of a useful map are distilled into concrete criteria. The ways that hybrid maps can address these criteria provide justification for producing them as an alternative to the current methodologies.

- **Multi-modal mapping processing pipeline** A mapping methodology was developed which selects the best sensor data for each map location from a redundant data set while respecting the inherent characteristics of each sensor. The individual steps in this processing pipeline were designed to address the map quality criteria.

4.3 Limitations and Future Work

4.3.1 Navigation

There are several lingering problems with navigation processing. Some remaining navigation error was apparent after refinement which indicates that the refinement process needs further improvement. The error was mainly obvious at the edges of objects. This could be the result of poorly understood covariances between submap registration. These covariances encode the relative weights of the various constraints acting on each node, and if one constraint is overvalued, it can pull the corresponding node and mapping sensor measurement out of alignment. More work is needed to determine why existing methods over-predict point cloud registration covariance. With noise in the constraints more accurately models, we can expect between alignment between maps.

4.3.2 Mapping more sites

Processing more data sets will certainly improve results for the mapping portion of the algorithm. Each new data set will reveal issues associated with its particular environment such a turbid water or high altitude. Methods to cope with each of these issues can be incorporated in turn into the mapping process without reformulating the existing algorithm.

4.3.3 Local versus global mapping

The proposed mapping method selects appropriate data based on local and neighborhood criteria. This is one class of solutions to this problem. A strength of this is that the tunable parameters have very obvious physical meanings and effects. Another class of solutions to the multi-modal mapping problem would be more unified or global solutions such as representing the entire surface with a spline fit or GP. These algorithms require good understanding of the error characteristics of the point clouds, and may in fact benefit from some of the proposed mapping techniques presented in this thesis. These unified approaches are powerful tools that offer an alternative to the proposed method.

4.3.4 Ground truth for navigation and mapping

Ground truth is absent from this thesis because it is difficult to obtain. Ground truth navigation data can be obtained using an LBL system. This data will available during the summer of 2013 so that the navigation algorithm can be compared against an accurate ground truth. Ground truth mapping data requires construction of a synthetic sea floor which can be measured using sensors other than multibeam and sonar (such as a kinect or laser scanner) with accurate navigation data. This is the ultimate ground truth data set to help evaluate and improve this algorithm.

List of Acronyms

SLAM Simultaneous Localization and Mapping

ROV Remotly Operated Vehicle

AUV Autonomous Underwater Vehicle

DVL Doppler Velocity Log

EKF Extended Kalman Filter

iSAM incremental Smoothing and Mapping

SAM Smoothing and Mapping

SEIF Sparse Extended Information Filter

MAP Maxiumum a Posteriori

SFM Structure from Motion

LBL Long Baseline

USBL Ultra Short Baseline

ICP Iterative Closest Point

DOF Degrees of Freedom

FOV Field of View

SSD Some of Squared Differences

GP Gaussian Process

BIBLIOGRAPHY

Agarwal, S., Snavely, N., Simon, I., Seitz, S. M., and Szeliski, R., "Building rome in a day," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 72–79.

Ballard, R., Stager, L., Master, D., Yoerger, D., Mindell, D., Whitcomb, L., Singh, H., and Piechota, D., "Iron age shipwrecks in deep water off Ashkelon, Israel," *American Journal of Archeology*, vol. 106, no. 2, April 2002.

Barkby, S., Williams, S., Pizarro, O., and Jakuba, M., "An efficient approach to bathymetric slam," in *2009 IEEE/RSJ International Conference on Intelligent robots and systems, Proceedings on*. Piscataway, NJ, USA: IEEE Press, 2009, pp. 219–224.

Barkby, S., Williams, S., Pizarro, O., and Jakuba, M., "Incorporating prior maps with bathymetric distributed particle slam for improved auv navigation and mapping," in *Proceedings of OCEANS 2009, MTS/IEEE Biloxi*. IEEE Press, Oct. 2009, pp. 1–7.

Barkby, S., Williams, S. B., Pizarro, O., and Jakuba, M. V., "A featureless approach to efficient bathymetric slam using distributed particle mapping," *Journal of Field Robotics*, vol. 28, no. 1, pp. 19–39, 2011.

Besl, P. and McKay, N., "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

Bradski, G. and Kaehler, A., *Learning OpenCV: Computer Vision with the OpenCV Library*, 1st ed. O'Reilly Media Inc., 2008.

Brandou, V., Allais, A. G., Perrier, M., Malis, E., Rives, P., Sarrazin, J., and Sarradin, P. M., "3D reconstruction of natural underwater scenes using the stereovision system IRIS," in *IEEE OCEANS '07*, Aberdeen, 2007, pp. 1–6.

Burt, P. J. and Adelson, E. H., "A multiresolution spline with application to image mosaics," *ACM Trans. Graph.*, vol. 2, no. 4, pp. 217–236, Oct. 1983. [Online]. Available: http://doi.acm.org/10.1145/245.247

Calder, B. R. and Mayer, L. A., "Automatic processing of high-rate, high-density multibeam echosounder data," *Geochemistry, Geophysics, Geosystems*, vol. 4, no. 6, pp. n/a–n/a, 2003. [Online]. Available: http://dx.doi.org/10.1029/2002GC000486

Cignoni, P., Rocchini, C., and Scopigno, R., "Metro: measuring error on simplified surfaces," in *Computer Graphics Forum*, vol. 17, no. 2. Wiley Online Library, 1998, pp. 167–174.

Curless, B. and Levoy, M., "A volumetric method for building complex models from range images," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM, 1996, pp. 303–312.

Dellaert, F. and Kaess, M., "Square Root SAM: Simultaneous localization and mapping via square root information smoothing," *Intl. J. of Robotics Research (IJRR)*, vol. 25, no. 12, pp. 1181–1204, Dec 2006.

Dellaert, F., "Factor graphs and gtsam: A hands-on introduction," 2012.

Douillard, B., Fox, D., Ramos, F., and Durrant-Whyte, H., "Classification and semantic mapping of urban environments," *The International Journal of Robotics Research*, vol. 30, no. 1, pp. 5–32, January 2011.

Eustice, R. M., "Large-area visually augmented navigation for autonomous underwater vehicles," Ph.D. dissertation, MIT/WHOI Joint Program, 2005.

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

Fallon, M. F., Kaess, M., Johannsson, H., and Leonard, J. J., "Efficient AUV navigation fusing acoustic ranging and side-scan sonar," in *2011 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, May 2011, pp. 2398–2405.

Gurram, P., Rhody, H., Kerekes, J., Lach, S., and Saber, E., "3d scene reconstruction through a fusion of passive video and lidar imagery," in *Applied Imagery Pattern Recognition Workshop, 2007. AIPR 2007. 36th IEEE*. IEEE, 2007, pp. 133–138.

Hover, F., Eustice, R., Kim, A., Englot, B., Johannsson, H., Kaess, M., and Leonard, J., "Advanced perception, navigation and planning for autonomous in-water ship hull inspection," *Intl. J. of Robotics Research, IJRR*, vol. 31, no. 12, pp. 1445–1464, Oct 2012.

Hurts, M., Cuf i Soler, X., and Salvi, J., "Integration of optical and acoustic sensors for 3d underwater scene reconstruction." *Instrumentation ViewPoint*, no. 8, pp. 43–, 2009. [Online]. Available: http://dialnet.unirioja.es/servlet/articulo?codigo=3201922

Jakuba, M. and Yoerger, D., "Autonomous search for hydrothermal vent fields with occupancy grid maps," in *Proc. of ACRA*, vol. 8, 2008, p. 2008.

Jakuba, M. V., Roman, C. N., Singh, H., Murphy, C., Kunz, C., Willis, C., Sato, T., and Sohn, R. A., "Long-baseline acoustic navigation for under-ice autonomous underwater vehicle operations," *Journal of Field Robotics*, vol. 25, no. 11-12, pp. 861–879, 2008. [Online]. Available: http://dx.doi.org/10.1002/rob.20250

Johnson-Roberson, M., Kumar, S., Pizarro, O., and Willams, S., "Stereoscopic imaging for coral segmentation and classification," in *IEEE OCEANS '06*, Sept 2006, pp. 1–6.

Johnson-Roberson, M., Pizarro, O., Williams, S., and Mahon, I., "Generation and visualization of large scale 3D reconstructions from underwater robotic surveys," *Journal of Field Robotics*, 2009 (in press).

Kenny, A., Cato, I., Desprez, M., Fader, G., Schüttenhelm, R., and Side, J., "An overview of seabed-mapping technologies in the context of marine habitat classification," *ICES Journal of Marine Science: Journal du Conseil*, vol. 60, no. 2, pp. 411–418, 2003.

Kinsey, J. C. and Whitcomb, L. L., "Preliminary field experience with the dvlnav integrated navigation system for manned and unmanned submersibles," In: Proceedings of the 1st IFAC Workshop on Guidance and Control of Underwater Vehicles, GCUV 03, Tech. Rep., 2003.

Kostylev, V. E., Todd, B. J., Fader, G. B. J., Courtney, R. C., Cameron, G. D. M., and Pickrill, R. A., "Benthic habitat mapping on the Scotian Shelf based on multibeam bathymetry, surficial geology and sea floor photographs," *Marine Ecology Progress Series*, vol. 219, pp. 121–137, Sep 2001.

Kunz, C., "Autonomous underwater vehicle navigation and mapping in dynamic, unstructured environments," Ph.D. dissertation, MIT-WHOI Joint Program, November 2011.

Lowe, D., "Object recognition from scale invariant feature descriptors," *Computer Vision, IEEE Conference on*, p. 1150, 1999.

Mahon, I., Pizarro, O., Johnson-Roberson, M., Friedman, A., Williams, S., and Henderson, J., "Reconstructing pavlopetri: Mapping the world's oldest submerged town using stereo-vision," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, 2011, pp. 2315–2321.

Mahon, I., Williams, S., Pizarro, O., and Johnson-Roberson, M., "Efficient view-based SLAM using visual loop closures," *Robotics, IEEE Transactions on*, vol. 24, no. 5, pp. 1002–1014, Oct. 2008.

Mayer, L. A., "Frontiers in seafloor mapping and visualization," *Marine Geophysical Researches*, vol. 27, no. 1, pp. 7–17, 2006.

Negahdaripour, S. and Firoozfam, P., "An ROV stereovision system for ship hull inspection," pp. 551–564, 2006.

Negahdaripour, S., Sekkati, H., and Pirsiavash, H., "Opti-acoustic stereo imaging: on system calibration and 3-d target reconstruction," *Trans. Img. Proc.*, vol. 18, no. 6, pp. 1203–1214, June 2009. [Online]. Available: http://dx.doi.org/10.1109/TIP.2009.2013081

Nicosevici, T., Gracias, N., Negahdaripour, S., and Garcia, R., "Efficient three-dimensional scene modeling and mosaicing," *Journal of Field Robotics*, vol. 26, no. 10, pp. 759–788, 2009. [Online]. Available: http://dx.doi.org/10.1002/rob.20305

Pizarro, O., Eustice, R., and Singh, H., "Large area 3-d reconstructions from underwater optical surveys," *Oceanic Engineering, IEEE Journal of*, vol. 34, no. 2, pp. 150–169, 2009.

Pizarro, O. and Singh, H., "Toward large-area mosaicing for underwater scientific applications," *IEEE Journal of Oceanic Engineering*, vol. 28, no. 4, pp. 651–672, October 2003.

Rigby, P., Pizarro, O., and Williams, S., "Towards geo-referenced auv navigation through fusion of usbl and dvl measurements," in *OCEANS 2006*, 2006, pp. 1–6.

Roman, C. and Singh, H., "Improved vehicle based multibeam bathymetry using submaps and slam," in *IROS'05: Proceedings of the 2005 IEEE/RSJ international conference on Intelligent robots and systems*, 2005, pp. 3662–3669.

Roman, C. and Singh, H., "A Self-Consistent bathymetric mapping algorithm," *Journal of Field Robotics*, vol. 24, no. 1-2, pp. 23–50, 2007.

Roman, C. N., "Self consistent bathymetric mapping from robotic vehicles in the deep ocean," Ph.D. dissertation, MIT/WHOI Joint Program, 2005.

Scharstein, D. and Szeliski, R., "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, pp. 7–42, April 2002. [Online]. Available: http://portal.acm.org/citation.cfm?id=598429.598475

Sedlazeck, A., Koser, K., and Koch, R., "3d reconstruction based on underwater video from rov kiel 6000 considering underwater imaging conditions," in *OCEANS 2009-EUROPE*. IEEE, 2009, pp. 1–10.

Sigurdsson, H., Carey, S., Alexandri, M., Vougioukalakis, G., Croff, K., Roman, C., Sakellariou, D., Anagnostou, C., Rousakis, G., Ioakim, C., Goguo, A., Ballas, D., Misaridis, T., and Nomikou, P., "Marine investigations of greece's santorini volcanic field," *Eos, Transactions American Geophysical Union*, vol. 87, no. 34, pp. 337–342, 2006. [Online]. Available: http://dx.doi.org/10.1029/2006EO340001

Singh, H., Howland, J., Yoerger, D., and Whitcomb, L., "Quantitative photomosaicking of underwater imagery," in *OCEANS '98 Conference Proceedings*, vol. 1, 1998, pp. 263–266 vol.1.

Singh, H., Roman, C., Whitcomb, L., and Yoerger, D., "Advances in fusion of high resolution underwater optical and acoustic data," in *Underwater Technology, 2000. UT 00. Proceedings of the 2000 International Symposium on*, 2000, pp. 206–211.

Singh, H., Whitcomb, L., Yoerger, D., and Pizarro, O., "Microbathymetric mapping from underwater vehicles in the deep ocean," *Computer Vision and Image Understanding*, vol. 79, no. 1, pp. 143–161, 2000.

Singh, H., Adams, J., Mindell, D., and Foley, B., "Imaging underwater for archaeology," *Journal of Field Archaeology*, vol. 27, no. 3, pp. 319–328, 2000.

Smith, R., Self, M., and Cheeseman, P., "Estimating uncertain spatial relationships in robotics," *Proceedings of the Second Annual Conference on Uncertainty in Artificial Intelligence*, pp. 167–193, 1986.

Thrun, S., Burgard, W., and Fox, D., *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. The MIT Press, September 2005.

Treibitz, T., Schechner, Y., Kunz, C., and Singh, H., "Flat refractive geometry," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, May 2011, PMID: 21576744. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/21576744

Triggs, B., McLauchlan, P. F., Hartley, R. I., and Fitzgibbon, A. W., "Bundle adjustmenta modern synthesis," in *Vision algorithms: theory and practice*. Springer, 2000, pp. 298–372.

Vasudevan, S., Ramos, F., Nettleton, E., and Durrant-Whyte, H., "Non-stationary dependent gaussian processes for data fusion in large-scale terrain modeling," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, 2011, pp. 1875–1882.

Vaughn, I., "Microbathymetry using self-contained navigation and simultaneous localization and mapping," Master's thesis, University of Rhode Island, 2012.

Williams, S. and Mahon, I., "Simultaneous localisation and mapping on the Great Barrier Reef," in *Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on*, vol. 2, May 2004, pp. 1771 – 1776 Vol.2.

Williams, S., Pizarro, O., Mahon, I., and Johnson-Roberson, M., "Simultaneous localisation and mapping and dense stereoscopic seafloor reconstruction using an auv," in *Experimental Robotics*, ser. Springer Tracts in Advanced Robotics, Khatib, O., Kumar, V., and Pappas, G., Eds. Springer Berlin / Heidelberg, 2009, vol. 54, pp. 407–416.

Yoerger, D. R., Jakuba, M., Bradley, A. M., and Bingham, B., "Techniques for deep sea near bottom survey using an autonomous underwater vehicle," *The International Journal of Robotics Research*, vol. 26, no. 1, pp. 41–54, 2007.

Yoerger, D., Bradley, A., Walden, B., Cormier, M.-H., and Ryan, W., "Fine-scale seafloor survey in rugged deep-ocean terrain with an autonomous robot," in *Robotics and Automation, IEEE International Conference on*, vol. 2, 2000, pp. 1787–1792.