Sonar Constrained Stereo Correspondence for Three-Dimensional Seafloor Reconstruction

Gabrielle Inglis
Department of Ocean Engineering
University of Rhode Island Narragansett, RI 02882
Email: gabrielle.inglis@gmail.com

Chris Roman
Graduate School of Oceanography
University of Rhode Island Narragansett, RI 02882
Email: cnr@gso.uri.edu

***Abstract*—There is a persistent need in the oceanographic community for accurate three dimensional reconstructions of seafloor structures. To meet this need underwater mapping techniques have expanded to include the use of stereo vision and high frequency multibeam sonar for mapping scenes 10's to 100's of square meters in size. Both techniques have relative advantages and disadvantages that depend on the task at hand and the desired accuracy. In this paper, we develop a method to constrain the often problematic stereo correspondence search to small sections of the image that correspond to estimated ranges along the epipolar lines calculated from coregistered multibeam sonar microbathymetry. This approach can be applied to both sparse feature-based and dense area-based stereo correspondence techniques. Data were collected on an underwater vehicle survey using a calibrated stereo rig and a multibeam sonar gathering coincident datasets. Overall, the constrained correspondence method shows improvements in the number and reliability of correct matches and allows for reduction in complexity of feature descriptors but it is reliant on the quality of both intrinsic and extrinsic sensor calibration.**

I. Introduction

In the oceanographic research and industrial communities accurate 3D visual and topographic reconstructions of the seafloor provide a wealth a knowledge. Both optical and acoustic mapping techniques are used for archeology [1], ship inspection [2], geology [3], [4], and biological habitat classification [5], [6]. Among the various acoustic and optical instrument configurations are stereo imagery and high frequency multibeam sonar, each of which has its own set limitations. In this paper we explore the potential of fusing data from the two sensing modalities to exploit their respective strengths and attempt to provide additional constraints in the scene reconstruction process.

Multibeam sonar has been a ubiquitous solution for underwater acoustic mapping due to its wide coverage, versatile range and invariance to turbidity. Recently, numerous off the shelf high frequency instruments with potential resolutions to centimeter scales have become available for use on underwater vehicles. However, acoustic range data can be sensitive to the surface and volume scattering characteristics of the environment. Many underwater sites, are characterized by very soft bottom materials that do not present a strong acoustic impedance difference. This can result in poor range resolution of small scale features.

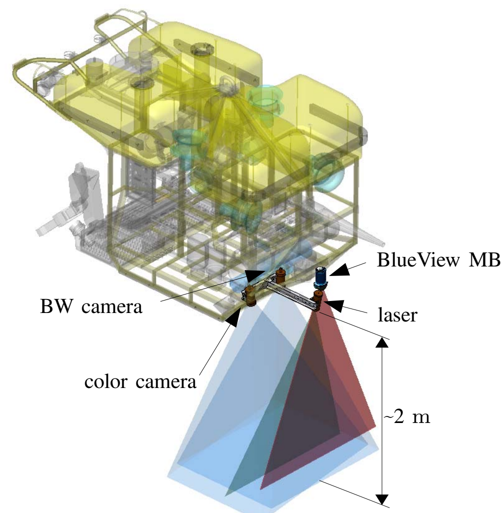


Fig. 1. The *Hercules* ROV with stereo cameras, green sheet laser and BlueView multibeam sonar. The Camera baseline is 0.425m. The geometry was designed for a nominal survey altitude of~3m.

An alternative to acoustic mapping is available in the form of stereo imaging [7]. Stereo vision uses multiple cameras to triangulate and produce a three dimensional reconstruction of the common viewed regions. Stereo imagery is not subject to the acoustic properties of the media, but has other inherent difficulties which are compounded in the underwater environment. In general, the geometry of cameras and multiple view configurations is well understood and detailed in several sources [8], [9]. The crux of stereo vision however, is establishing feature correspondences from which the triangulated points are calculated. This is the subject of much continued research [10], [11], [12]. Classical stereo correspondence methods

978-1-4244-5222-4/10/$26.00 ©2010 IEEE

tend to fail in low contrast, low texture, and occluded regions. In underwater imaging, these problems are compounded. Extraneous noise from backscatter, uneven lighting, low contrast and the generally unstructured nature of underwater scenes cause additional difficulties [13].

While stereo imaging is heavily researched in the land robotics community, very few algorithms specifically address the difficulties associated with underwater applications. The low contrast nature of the images makes it difficult to obtain feature encryption that is unique across the entire image, making a one to all inter-image correspondence search method impractical due to the high numbers of false matches. This difficulty has been most successfully addressed by constraining the correspondence search to small regions along related epipolar lines using an approximation of the scene depth [14], [15]. This reduces the region across which the feature encryption must be unique and in total reduces the likelihood of false matches. Eustice and Pizarro used a planar scene assumption and inter image epipolar geometry determined from the vehicles navigation data [14], [15] to limit feature matching. Lanser and Lengauer take a similar approach using a significantly more complex estimate of the scene based on a CAD model [11].

In this paper we adapt the range constraint concept and expand it to utilize the more detailed information provided by coincident multibeam sonar data. Section II outlines an approach for constraining both sparse feature matching and dense stereo disparity estimation with local bathymetry data over the imaging area. Our goal is to reduce the correspondence search to a small region of uncertainty dictated by the accuracy of the multibeam bathymetry. Section III presents results for two surveys and Section IV discusses the potential benefits and limitations of the approach.

II. Methods

The multisensor approach to stereo matching begins with the acquisition of multibeam sonar data, stereo images and navigation data from an underwater vehicle seafloor survey. This data is preprocessed to reduce error and the multibeam sonar and stereo data sets are mapped into a common coordinate system using the navigation data and then the features matches are established. Feature correspondences can then be triangulated into three dimensions, and finally post-processing is undertaken to smooth and texture map the data, creating the final scene model.

A. Survey Methods & Preprocessing

1) Instrumentation: The data for this work was gathered during vehicle based archaeological surveys in August 2009 using the Institute for Exploration remotely

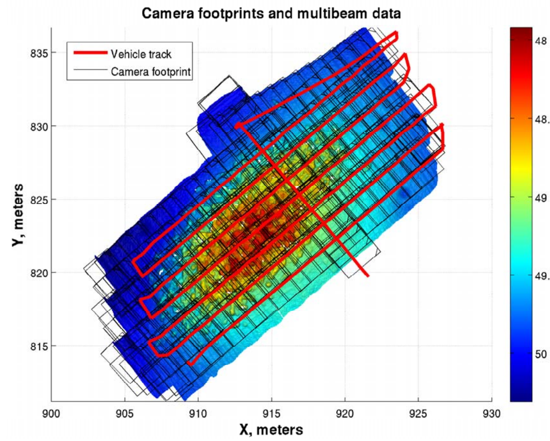


Fig. 2. A sample vehicle survey to acquire multibeam sonar ranges and overlapping image footprints at an archaeological site in the Aegean Sea. The background is the micro-bathymetry map created using the multibeam sonar.

operated vehicle (ROV) *Hercules* (fig. 1). Serpentine surveys $\mathcal{O}(100\text{m}^2)$ were to gather simultaneous acoustic and optical imagery (fig. 2). An outline of the survey instruments used on the *Hercules* ROV is presented in Table I.

Acoustic data was collected using a Blueview MB2250 Multibeam sonar. This is a 2.25 MHz system with a $45°$ field of view. Images were taken using a rigid stereo rig fitted with two Prosilica EC1380 cameras. These were mounted 425mm apart within pressure housings with glass viewports. Their optical axes were verged to intersect at 3m, the average anticipated survey altitude. The artificial lighting came from two Ocean Imaging Systems model M3831 flashbulbs which were hardware triggered off of the master camera. These were mounted on the forward half of the vehicle to minimize the common volume of water imaged by the camera and the two strobes.

2) Calibration: A precise characterization of the cameras' projective properties must be obtained in order to derive any quantitative information, such as a 3D reconstruction, from the images. The perspective projection matrix transforms points in the world coordinate system into their projections on the image plane. $\mathtt{P}$ is the perspective projection matrix containing intrinsic and extrinsic parameters of the camera which govern the mapping

$$\underline{\mathbf{u}} = \mathtt{P}\underline{\mathbf{M}} \tag{1}$$

where $\underline{\mathbf{u}} = [\mathbf{u}, 1]^\mathsf{T}$ is the homogeneous representation of the pixel coordinates $\mathbf{u}$, and $\underline{\mathbf{M}}$ is the homogeneous representation of the three dimensional point $\mathbf{M}$. The calibration information is gathered through a calibration process. Images of a known checkered target are entered into the Matlab Camera Calibration Toolbox created by

TABLE I
Hercules VEHICLE INSTRUMENTATION

	Function	Instrument
Navigation	Depth Sensor	ParoScientific Pressure Sensor
	Compass and Attitude Sensor	Octans Fiber Optic Gyro
	Doppler Velocity Log	Teledyne RDI
Optics	Still Cameras	Prosilica EC1380 monochrome
		Prosilica EC1380C Bayer color
Acoustics	Multibeam Sonar	BlueView MB2250 (2009)

Bouguet [16], which determines the intrinsic properties of the camera, specifically those related to lens distortion and projection.

Calibration also determines the relationship between the cameras which is used to triangulate the final feature correspondences. From this relationship, a point transfer mapping can be derived which relates the location of matched features in the two image planes to corresponding range of the actual object in the camera centered 3D coordinate system:

$$\mathbf{u}' = \frac{\mathtt{H}_\infty \underline{\mathbf{u}} + \mathtt{K}\mathbf{t}/Z}{\mathbf{H}_\infty^{3\mathsf{T}} + t_z/Z} \tag{2}$$

where $\mathtt{K}$ constrains the projective properties of the camera, $\mathtt{H}_\infty = \mathtt{KRK}^{-1}$ is the infinite homography, $mathttR$ and t are the rigid transformation between cameras, $\mathbf{H}_\infty^{3\mathsf{T}}$ is the third row of $\mathtt{H}_\infty$ and t_z is the z component of the translation $\mathbf{t}$. If all of the variables are precisely known, then Equation 2 is an exact mapping from a point in the left view, i, to the same point as it appears in the right view, j. However, if each calibration parameter is modeled as a random variable with a known mean and variance, Equation 2 results in a region that is likely to constrain the point

3) Image Preprocessing: Preprocessing the images can reduce artifacts which cause errors during the matching process. Uneven lighting patterns can be modeled by averaging all of the images from a single survey to produce a mask. Multiplying the mask by the image balances out the light field [17]. Radial lens distortion is removed using the distortion coefficients estimated during the calibration process [18]. Contrast is increased using adaptive histogram equalization (fig. 3).

The right hand camera is a color camera with a Bayer patterned CCD. The difference between the black and white and color images due to Bayer demosaicking is a blurring effect analogous to representing the same features at slightly different spatial scales. To compensate for this, the black and white image is blurred using a Gaussian filter [19]. Finally the images are converted from 16 bit to 8 bit representations which reduces computational burden during matching as well as masking noise.

B. Data Coregistration and Backprojection Considerations

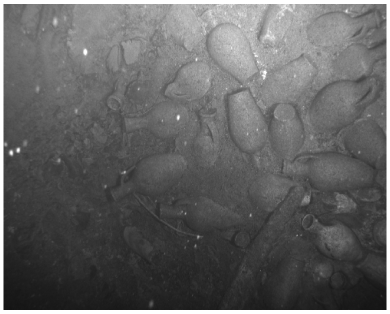
After preprocessing, the imagery and the multibeam sonar data must be expressed in a common coordinate system to be used during matching. The correspondence search occurs within the image coordinate system making bathymetric data most practical when represented in this coordinate system where it is easy to associate each sonar range estimate with a photometric feature. We use several methods to backproject terrain data, but all revolve around the basic mechanics of mapping points from the sonar scene model into the camera reference frame and then essentially 'taking a picture' of the points or projecting them onto the image plane. This is done by transforming the bathymetry into the camera's reference frame via sensor offsets and then into the image frame via the camera projection matrix from Equation 1

1) Sensor Registration: The spatial relationship between sensors is used to transform sensor data into common coordinate systems. At first, the sensor offsets are measured by hand. This gives an initial estimate for the sensor offsets which is accurate within several cm or degrees. Therefore, we endeavor to algorithmically refine the offsets. In this work, the Iterated Closest Point (ICP) algorithm is used to align point clouds from each sensor to come up with the best transformation relating the sonar to the cameras. After the offset measurements have been refined, the offset transformation can be considered static.

2) Backprojection: Once the sonar data is in the camera reference frame, it can be transformed into the image place via the camera projection matrix. Certain areas of the seafloor present significant relief which results in occlusion during backprojection. This is remedied by occlusion culling using a z-buffering algorithm [20].

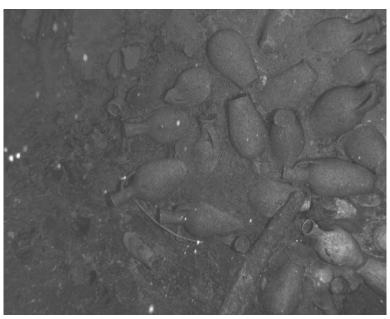
The first method for using backprojected sonar scene models is intended for sparse correspondence searches. Each point in the sonar scene model is tagged with its range and simply projected onto the image plane according to Equation 1 and is tagged with its original range creating a (u, v, Z) triplet. A given range is now be expressed in terms of its image plane pixel coordinates.

Often with dense stereo, computations can be sped



(a) Raw Image

(b) Lighting Model



(c) Light Balanced Image

(d) Adaptive Histogram Equalized Image

Fig. 3. Preprocessing Steps. The goal of preprocessing is the enhancement of the signal to noise ratio, which allows for more unique and robust feature matches to be established.

up by rectifying the images which reduces the search to a 1D pixel scanline [21]. Rectification is the process of rotating the cameras around their optical center and altering focal lengths such that conjugate epipolar lines become colinear and parallel to the horizontal image axis. This is accomplished using the algorithm developed by Fusiello et al [22] for rectification of a calibrated stereo rig of unconstrained geometry. The result is a new perspective projection matrix $\mathtt{P}_{rectified}$ for each camera.

Points are backprojected via the new projection matrix and range tags are converted into disparity which can be related to range for a rectified rig using the camera focal length, f, and the rig baseline, b:

$$d = \frac{bf}{r} \tag{3}$$

This is a convenient form for expression of range data for the dense correspondence search.

We expressed the dense disparity estimate using several different approximations, each imposing a different level of constraint on the search. The simplest method is to determine the image-wide minimum and maximum disparity estimate from the backprojected sonar scene model (figs. 4(c, f)). This is consistent with many dense correspondence algorithms which require parameterization using a globally constrained disparity search range.

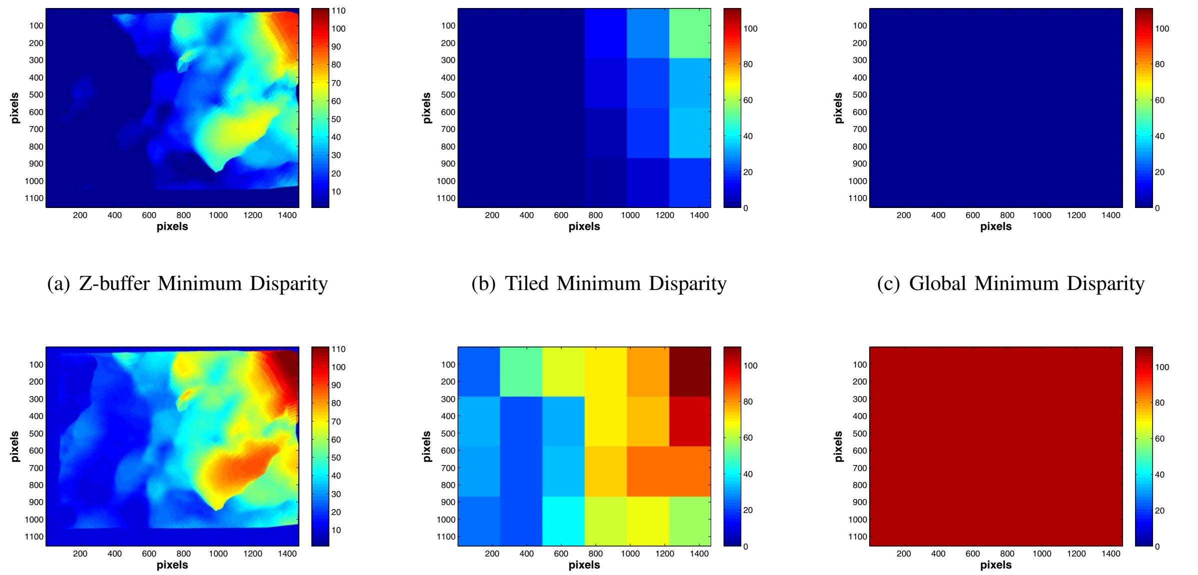
The next simplest method is to divide the backprojected image into a number of tiles and obtain the minimum and maximum predicted disparity within each tile from the backprojected bathymetry (figs. 4(b, e)).

The greatest amount of constraint comes from the most detailed disparity estimate. This uses every sonar range that backprojects onto the image plane. The sonar scene map is sparse, so it is tessalated using Delaunay triangulation and the vertices are tagged with their ranges. Then the vertices are back projected using using $\mathtt{P}_{rectified}$. Pixels on in the disparity estimate falling within each triangle are labeled with the mean range of each of the three vertices. This approach was used–rather than interpolation of range across the image–because it can be effectively vectorized in Matlab for accelerated processing (figs. 4(a, c)).

C. Terrain Constrained Correspondence

Once the terrain model has been transformed into a useful coordinate system via sensor offsets and backprojection, the correspondence search can be initiated. The remainder of this section focuses on the feature matching. It outlines the approach to bathymetry constrained stereo correspondence in both the sparse and dense search frameworks.

1) Sparse Methods: Sparse stereo correspondence first requires the extraction and description of a sparse set of interest points from the image, followed by a search



(a) Z-buffer Minimum Disparity (b) Tiled Minimum Disparity (c) Global Minimum Disparity

(d) Z-buffer Maximum Disparity (e) Tiled Maximum Disparity (f) Global Maxmum Disparity

Fig. 4. Disparity estimates derived from backprojected bathymetry can be used as a z-buffered projection (a), (d); tiled (b), (e); or as a global approximation (c), (f).

for interimage matches which maximize a similarity measure.

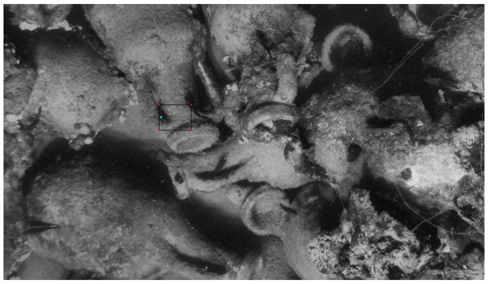
Feature detection consists of determining which points in an image represent ‘interesting’ pieces of information. Because features form the main primitives of many image processing algorithms, the quality and proper selection of the feature often determines the success of the algorithm. A feature must be detected identically in both images for the possibly of a correct match to exist, consequently sparse terrain constrained correspondence relies as heavily on feature detection repeatability as any other sparse correspondence algorithm. However, scene model constraints can reduce uniqueness demands on the feature *description* from global to local.

This work utilizes ~$30,000$ features in each image encoded as the robust Scale Invariant Feature Transform (SIFT feature). The matching success of these features is compared with that of the less robust image patch descriptor in order to evaluate the ability of the constrained algorithm to perform in the absence of rich feature descriptor. SIFT features can be compared with one another by finding the Euclidean distance between feature descriptors. Images patches are compared using NCC, a common alternative to Sum of Squared Differences (SSD) in dense correspondence algorithms.

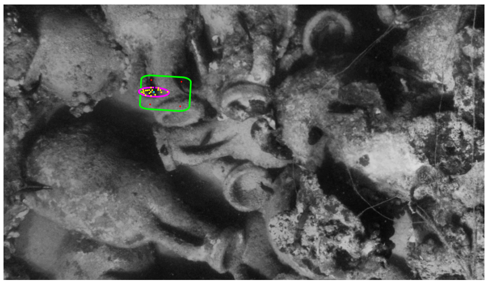
The correspondence algorithm implementation is derived from that of Eustice [23]. The processing requires calculations which routinely exceed the limitations of the RAM available to Matlab. Therefore, the key image is broken down into tiles with with an average of 300 features in each tile. The bounds of this tile are transformed into the opposing image using the point transfer function with uncertainty with the mean range across the tile as the point transfer range Z and the variance of Z to create a 95% confidence interval around the mapped tile Equation 2. The putative correspondence set is established for the tile between view i and j and and then the roles are reversed and matches are established between j and i. The match needs to be the maximum similarity score in both views in order to be considered. Then the second best similarity score is found. If the best similarity score is two times that of the second best, then the match is kept. This helps avoid mismatches, especially in low texture areas. After correspondences have been established in both directions and have met uniqueness requirements, the algorithm progresses to the next tile (fig. 5).

2) Dense Methods: Dense correspondence searches attempt to find stereo matches for each pixel in the image, regardless of photometric interest level. The common block matching algorithm is parameterized by a minimum-maximum disparity estimate that is uniform across the image [24]. In this work, the search is constrained further by setting a minimum and maximum disparity search limit based on sonar range data that varies across the image and thus employs stricter constraints than a globally constant disparity bound(fig. 6).

After image rectification, and terrain back-projection



(a)



(b)

Fig. 5. Tiled sparse correspondence search. The red points are the vertices of the tile, the green outline is the tile mapped from left to right with uncertainty. The point in the left image is compared with the features that fall within the ellipse in the right image.

via $\mathtt{P}_{rectified}$, the disparity map is approximated using one of the methods mentioned previously and keyed to the left hand image. Then a block matching algorithm is used to actually perform the stereo correspondence search. This method is similar to the block matching employed in OpenCV [21]. Matches are selected at the pixel pair with maximum normalized cross correlation (NCC)(fig. 6).

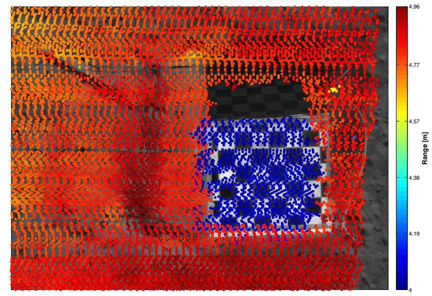
A significant concern in dense correspondence algorithms is that the parameterization of the algorithm usually involves significant supervision. Parameters that require attention can include window size, texture thresholds, and uniqueness thresholds. Terrain knowledge could conceivably determine these parameters increasing the autonomy of the algorithm.

III. Results

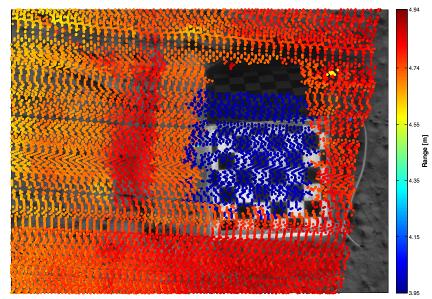
A. Backprojection of Terrain

The first attempt at refining sensor offsets using algorithmic means attempts to match reconstructions of natural scenes captured using multibeam and stereo cameras. The algorithm failed to converge at a solution. Instead it continued to drift away from the initial estimate for as many iterations as it was allowed to run.

The second trial attempted to align corresponding multibeam and stereo point clouds acquired over the man-mad target. Again, convergence was unachievable. Figure 7(a), show the best manual alignment that was achieved, compared with the tenth iteration of ICP (fig. 7(b)) after which the offset began to diverge considerably. Ultimately a manual alignment was used to determine the sensor offsets.



(a) Manual Alignment



(b) ICP Alignment at 10th iteration

Fig. 7. [Manual refinement of transformation between sonar point cloud and image data compared with ICP sensor alignment. Better results are currently achieved using manual alignment.

B. Sparse Correspondence Results

Simply using a one to all correspondence search based on matching SIFT features established a baseline. The the triangulated results of this exhaustive correspondence search are displayed in Figure 8(a). Typically, fewer correct matches resulted from this method than from the constrained methods.

There were two different sonar constrained searches. The first one used SIFT features. The resulting triangulated matches are shown in Figure 8(b). More than twice as many matches made it through the outlier rejection step for Harris corners and the triangulated results are displayed in Figure 8(c) and Table II.

Harris corners resulted in more correct matches in textured areas than the SIFT features. While both types of features had fewer correct matches in the untextured areas, SIFT was clearly the better feature for matching in these areas.

In all cases, the correct matches are focused on the surface of the closest amphora. Fewer correct matches are apparent in the sandy areas where man-made objects

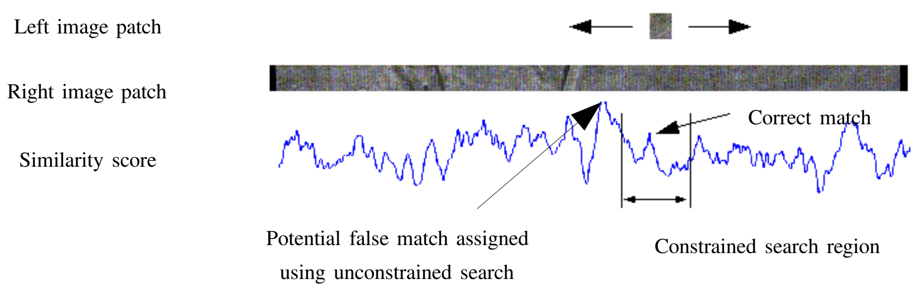
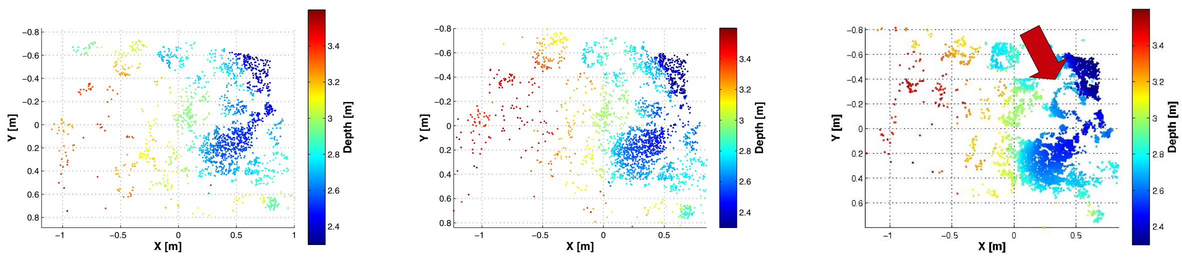


Fig. 6. Dense correspondence search. An image patch in the left image is compared with a corresponding strip in the right image, similarities are calculated for each possible match using normalized cross correlation. Note that in this case that the correct match is a local maximum that can only be correctly identified using a constrained search. An unconstrained search would have incorrectly identified the global maximum of the similarity measure.



(a) Unconstrained SIFT (b) Constrained SIFT (c) Constrained Harris

Fig. 8. Results of constrained and unconstrained sparse correspondences. SIFT features were extracted and matched in an exhaustive search (a) and a terrain constrained correspondence search (b). Harris features compared using terrain constrained normalized cross-correlation (c).

TABLE II
RESULTS OF FEATURE MATCHING IN ONE TYPICAL STEREO PAIR

Method	Number of Features	Number Matched	Number of Inliers
SIFT Unconstrained	L:22872 R:29913	2307	1952
SIFT Constrained	L:22872 R:29913	3154	2523
Harris Constrained	L:30000 R:30000	7612	5471

are sparse. And, while all matching efforts began with approximately the same number of features ($\sim$30,000 in each image), Harris features compared using normalized cross correlation resulted in the greatest number of correct matches.

C. Dense Correspondence Results

The images were rectified such that related epipolar lines become coincident with identical pixel scanlines, then the correspondence search is executed one pixel at a time. Figure 9 shows disparity maps for each of the three disparity approximation methods. Disparity is related to range though Equation 3, so these disparity maps are analogous to depth maps where warm colors represent closer objects and cool colors represent farther objects. The key image shown in Figure 9(a) is in the same coordinate system as the disparity maps and works as an adequate visual check for accuracy.

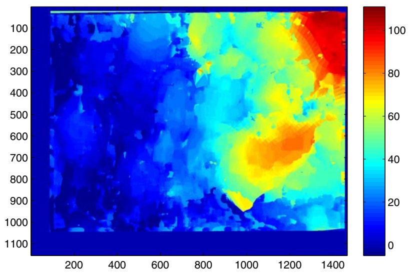
Each of the disparity maps show noise issues around the edges as well as in regions corresponding to low texture of the sandy bottom and shadows between amphora. The gridded pattern in Figure 9(b) on the curved surface of the amphora and the hard vertical line in the upper left corner of Figure 9(c).

IV. DISCUSSION

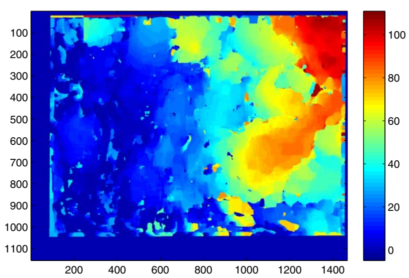
A. Registration and Backprojection

In attempting to co-register the multibeam and stereo datasets, the most successful method was manual measurement of the offsets. The ICP algorithm failed to converge on a rigid transformation for the sensor offsets. For natural scenes, it is possible that limited relief in the

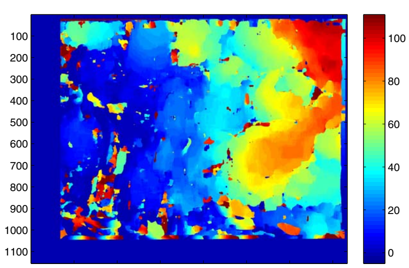
(a) Key (left) image



(b) Z-buffered search constraint



(c) Tiled search constraint



(d) Global min max constraint

Fig. 9. Disparity maps created using varying levels of terrain constraint. (b) uses the greatest degree of constraint followed by (c), (d).

reconstructions and sparsity and uneven distribution of stereo points contributed to the lack of convergence.

When ICP was used with the man-made target, several factors could have contributed to lack of convergence. The target had one angled face which was intended to give convergence a unique solution, instead of four possible solutions available when the three dimensional target is a cube. This angled face caused sonar multipath. Consequently the entire sonar representation of that face had to be cropped out leaving a large hole in the data. In the stereo data set, areas of sediment immediately surrounding the target were feature poor leading to very little reconstruction around the target. The sparsity of points surrounding the target probably caused the lack of convergence. Furthermore, since neither the stereo nor the multibeam was capable of characterizing the angled face, there were four possible convergence solutions instead of just one.

Additionally, there is one final obstacle to convergence. The sonar and stereo often had slightly different representations of the flat parts of the seafloor. This could be because stereo characterizes the immediate surface and its optical properties while sonar characterizes the acoustic reflective properties of the surface. The nature of this discrepancy is beyond the scope of this paper but warrants additional investigation.

B. Sparse Correspondence

More matches resulted from the constrained searches than from the unconstrained search. In constrained searches any given feature in one image is compared with several hundred features in the opposing image, instead of ~$30,000$. Therefore there are fewer false matches that might also have a high similarity score causing the correct match to be rejected altogether. Thus the likelihood of a maximum similarity score meeting the threshold of uniqueness increases and the reduced candidate correspondence pool increases the likelihood of this being the right match.

There are more matches for NCC than for SIFT descriptors. NCC is not strongly invariant to any transformations and is very likely to cause mismatches, however, it captures many aspects of the image area that are not captured by abstractions such as SIFT descriptors which represent hundreds of pixels as 128 values. The extra information available in NCC makes matches possible where they weren't possible before, and the addition of the constraint not only makes this computationally intensive method plausible by reducing the number of candidate correspondences, it also reduces the mismatches due to noise, making a previously difficult method more viable.

A greater number of correct matches appear in the feature poor regions using SIFT descriptors than NCC.

An image patch contains only image information local to a 40×40 pixel area. SIFT features can be extracted at multiple scales and can characterize wide areas. The regions in which SIFT features matched where there was very little local information were large scale SIFT features that characterized the relationship of structures around the low texture area. It is important to note, however that features extracted at large scales are by nature less well localized than small scale features.

Harris features compared using NCC had higher density in well textured areas than SIFT features matched using the same terrain constraints, for example the circular rim of the arrowed structure in Figure 8(c). A SIFT feature in the same region would not be as responsive since it abstracts the region into gradient of sections and depending on scale/orientation may not give enough weight to that feature of the structure.

C. Dense Correspondence

Dense correspondence produced a variety of results depending on the approximation of disparity used. The greatest degree of constraint creates the most visually appealing results. The matches are constrained to a small region around the sonar scene model, so in areas where the lack of unique texture would result in a mismatch, the mismatch is limited to near the sonar scene estimate.

Another characteristic of the more constrained method is the bathymetry artifacts that are present. These departures from the image structure follow the sonar terrain map. The discrepancies result from a combination misalignment between the data sets and and erroneous sonar returns. It is possible that a more complete understanding of these errors would allow us to incorporate a prediction of this effect and eliminate it.

Noise is most apparent in the least constrained correspondence searches. While minimal constraint can be useful simply to parameterize existing implementations of other matching algorithms such as Konolidge's Block Matching, it leaves the most possibility for mismatches to result in noise obscuring the scene structure. These mismatches are most frequent in areas of low signal to noise ratio, such as shadows and sandy areas.

Tiles also leave artifacts of the underlying disparity search constraints. In the upper right hand corner of Figure 9(c), there is a sharp vertical transition which corresponds to the edge of a tile. To the left of this line, the sonar scene model wasn't an adequate representation of the scene and the proper matches fell outside the constrained search window. To the right of this vertical line, the correct matches fell within the search constraints and were properly assigned. In this region, the z-buffer wasn't able to characterize this structure at all due to inadequate representation of the structure in the scene model, however the nearly unconstrained disparity map fully represents the structure.

D. Outlier Rejection

Outlier detection for constrained stereo matching is difficult, and often times, a few outliers remain. Least Median of Squares (LMedS) has proven to be the most successful outlier rejection scheme when applied to all three trials (one unconstrained, and two constrained), however it is more reliable at removing all outliers in the unconstrained case. This is because outliers here are just as likely to be geometrically distant from the model as they are to be close to the mode. However, for the constrained cases, bad matches were always geometrically proximal to the good matches because the distant potential bad matches had already been removed via the sonar constraint. LMedS did a decent job of removing those that didn't fit the model, but some noise remains.

Related methods include limiting triangulation error. This approach is very effective. It rejects fewer correct matches while retaining a similar number of false matches as LMedS. Another advantage is that restricting triangulation error is not iterative so it takes much less computation time. The main disadvantage of this method is that selecting the optimal upper limit of triangulation error requires some tuning, whereas LMedS, while conservative, calculates its own threshold using the Robust Sigma method described in Faugeras [9].

V. Conclusion

This paper has demonstrated a method for using sonar scene models to constrain a stereo correspondence search as well as some of the considerations necessary for this process. These considerations include rig geometry and calibration, sensor offset refinement, sparse and dense terrain backprojection and outlier rejection.

Establishing an accurate set of sensor offsets is critical to multimodal mapping, however this work concludes that approaching this problem through algorithmic means is quite challenging, due to a sparsity of structure on the seafloor and sparsity of multibeam and stereo reconstructions in general. Consequently, manual refinement of sensor offsets is still the most viable method. Despite the importance of correct sensor offsets, it is possible to continue using the backprojected terrain model even as offset accuracy degrades just by reducing the level of constraint which it imposes. Currently the best degree of constraint is selected using calibration uncertainty in the sparse case and experimentation in the dense case. Improving the quantification of error and expanding it to include uncertainty in sonar maps and sensor offsets will allow the level of constraint to be tuned in a principled manner.

Using the terrain constraint on sparse matching increased the number of correct matches. It also reduced uniqueness and noise robustness demands on the feature descriptors. Though very few mismatches remained in final reconstructions, occasional false matches did slip through the outlier rejection criteria. Selection of rejection criteria that doesn't eliminate many good matches is a difficult problem and finding a slightly more conservative set of rejection criteria would benefit the final texture mapped reconstructions.

The terrain constraint helps dense correspondence searches reject structures unrelated to the seafloor such as fish. Varying levels of constraint can be used to compensate for structures that are incorrectly represented in the multibeam, however the less constraint that is used, the noisier the final disparity map appears. Constrained correspondence has the added benefit of seamlessly substitutes the sonar scene model into the disparity map in the case of ambiguous matches. In its simplest form, it can be used to parameterize dense correspondence searches that have been developed by various researchers simply by making a minimum and maximum disparity estimate available. This method also provides an alternative to vision based initial disparity estimates which have been used in the past. It provides a regular grid of estimated disparity instead a set of sparse points concentrated in feature rich areas which provides less constraint in low texture areas where a stereo algorithm needs the most constraint.

There are several avenues of future research to improve this methodology. First of all, improved error analysis could allow for unsupervised selection of constraint level which reflects the confidence in the navigation, sonar reconstruction, sensor offsets as well as camera calibration and feature localization. The next step beyond this is to expand the image based reconstruction to a multiple image reconstruction that incorporates the site-wide sonar reconstruction as a constraint on the multiple image based reconstruction.

ACKNOWLEDGMENT

The authors would like to acknowledge The Institute for Exploration and the Archeological Oceanography 2009 Field Program.

REFERENCES

[1] 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.

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

[3] 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, 2000. Proceedings. ICRA '00. IEEE International Conference on*, vol. 2, 2000, pp. 1787–1792 vol.2.

[4] 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 *OCEANS 2007*, Aberdeen, 2007, pp. 1–6.

[5] 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. [Online]. Available: http://www.int-res.com/abstracts/meps/v219/p121-137/

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

[7] 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, July 2000.

[8] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.

[9] O. Faugeras, Q. Luong, and T. Papadopoulous, *The Geometry of Multiple Images: The Laws that Govern the Formation of Images of a Scene*. Cambridge, MA, USA: MIT Press, 2001.

[10] S. Fan and F. P. Ferrie, "Photo hull regularized stereo," *Image and Vision Computing*, vol. In Press, Corrected Proof, 2008.

[11] S. Lanser and T. Lengauer, "On the selection of candidates for point and line correspondences," in *In International Symposium on Computer Vision*. IEEE Computer Society Press, 1995, pp. 157–162.

[12] 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, 2002.

[13] J. Jaffe, K. Moore, J. McLean, and M. Strand, "Underwater imaging: Status and prospects," *Oceanography*, vol. 14, no. 3, pp. 66–76, 2001.

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

[15] 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.

[16] J.-Y. Bouguet, "Camera Calibration Toolbox for Matlab," retrieved March 2008. [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc/index.html

[17] 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.

[18] J. Heikkila and O. Silven, "A four-step camera calibration procedure with implicit image correction," in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, June 1997, pp. 1106–1112.

[19] W. Kim, A. Ansar, R. Steele, and R. Steinke, "Performance analysis and validation of a stereo vision system," in *Systems, Man and Cybernetics, 2005 IEEE International Conference on*, vol. 2, Oct. 2005, pp. 1409–1416.

[20] T. Theoharis, G. Papaioannou, and E.-A. Karabassi, "The magic of the z-buffer: A survey," in *Proc. of 9th International Conference on Computer Graphics, Visualization and Computer Vision, WSCG*, 2001.

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

[22] A. Fusiello, E. Trucco, and A. Verri, "A compact algorithm for rectification of stereo pairs," *Machine Vision and Applications*, vol. 12, no. 1, pp. 16–22, 2000.

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

[24] K. Konolidge and D. Beymer, "Sri small vision system user manual," May 2007.