

2009

Terrain Constrained Stereo Correspondence

Gabrielle Inglis

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

Chris Roman

University of Rhode Island, croman2@uri.edu

Follow this and additional works at: <https://digitalcommons.uri.edu/gsofacpubs>



Part of the [Engineering Commons](#), and the [Oceanography Commons](#)

Terms of Use

All rights reserved under copyright.

Citation/Publisher Attribution

Inglis, G., & Roman, C. (2009). Terrain constrained stereo correspondence. *OCEANS 2009, MTS/IEEE Biloxi - Marine Technology for Our Future: Global and Local Challenges*, pp.1-9, Oct 26-29.

Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5422106&isnumber=5422059>.

This Article is brought to you for free and open access by the Graduate School of Oceanography at DigitalCommons@URI. It has been accepted for inclusion in Graduate School of Oceanography Faculty Publications by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons@etal.uri.edu.

Terrain Constrained Stereo Correspondence

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 micro-bathymetry. 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 heavily reliant on the quality of the intrinsic and extrinsic calibration of the camera and sonar systems.

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, including areas of the Black Sea where the data presented here was collected, 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.

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 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 discusses the potential benefits and limitations of the approach.

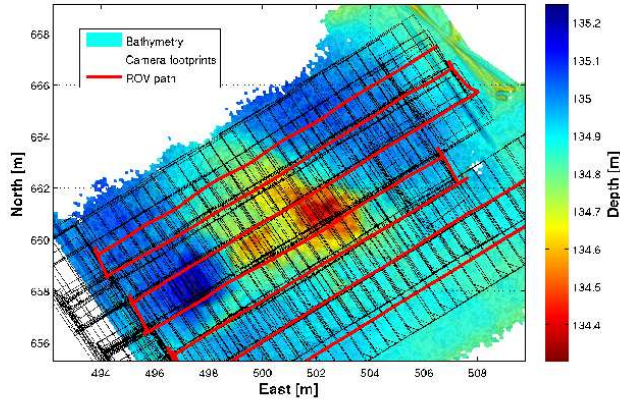


Fig. 1. A sample vehicle survey to acquire multibeam sonar ranges and overlapping image footprints at an archeological site in the Black Sea. The background is the micro-bathymetry map created using the multibeam sonar.

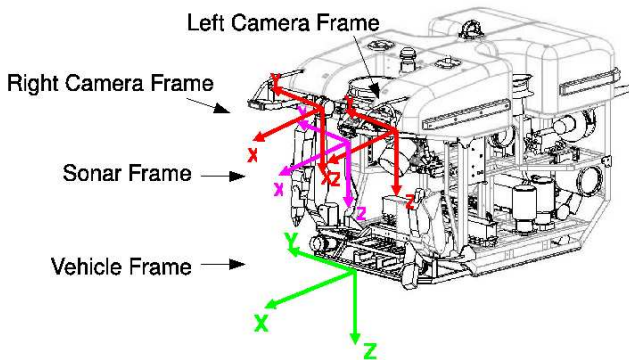


Fig. 2. Relevant sensor coordinate reference frames. Each sensor frame is related to the vehicle frame via a rigid motion transformation.

II. METHODS

A. Survey Methods & Preprocessing

The data for this work was gathered during vehicle based surveys in the Black Sea in August of 2007 using the Institute for Exploration ROV *Hercules*. The surveys took place at two shipwreck sites and were executed in a mowing-the-lawn pattern during which simultaneous multibeam and stereo data were gathered (fig. 1).

The sonar data was acquired with an Imagenex Delta-T 675kHz multibeam sonar. Images were taken with a rigid stereo rig fitted with two Proslica EC1380 cameras; one black and white and one Bayer patterned color CCD, mounted within pressure housings. The stereo rig was calibrated using the Matlab Camera Calibration Toolbox to establish the projective properties of the camera as well as the rigid transformation between cameras[16].

Before any analysis is performed on the images a number of preprocessing steps are taken. A random sample of images from each survey are averaged in order to model the lighting pattern of the vehicle mounted light source (fig. 3(b)). Each image is then corrected using this pattern to create an effective even illumination across the image. The Bayer filtered images normally have one quarter the resolution of black and white images. Therefore the black and white images are convolved with a Gaussian filter so that features can be extracted and correlations established on similar physical scales in each image. This step improves matching with less robust features such as sum of squared differences or normalized cross-correlation between windows[17]. Texture in underwater images are characteristically low in contrast, however texture is very useful in establishing correspondences, therefore the image textures were enhanced using adaptive histogram equalization (fig. 3(d)). Finally, we compensate for lens distortion using distortion coefficients established during the calibration procedure and the method described by Heikkila [18].

B. Point Cloud Registration

In order to use the multibeam and the stereo image data together, the two cameras and the sonar must be located in a common frame of reference. The sensor offsets can be initially estimated by manually measuring the sensor positions with respect to the preselected origin in the vehicle coordinate system. Small attitude biases in each instrument can be refined independently by enforcing a minimum range variance along and between tracklines while making incremental adjustments to the angular offsets[19]. In order to further refine the offsets for each instrument, a point cloud registration technique can be used to coregister the triangulated stereo from an unconstrained search and the processed multibeam range data, such as the Iterative Closest Point algorithm using Horn's method for determining relative orientation. With the location and attitude of both instruments established within the vehicle frame of reference, the multibeam ranges are transformed into the camera coordinates where they can be used to constrain the stereo correspondence search.

C. Terrain Constrained Correspondence

Previous work in underwater stereo has utilized planar approximations of the scene to guide stereo matching[14], [20]. This reduces the number of feature comparisons that must be made for each feature during an exhaustive search. When a more complex scene model is available, such as sonar bathymetry, it can be used to constrain the search to a greater degree. There are two stereo matching frameworks in which it is possible to apply this method. The sparse approach matches only image points which meet some photometric interest requirement, while dense stereo attempts to compute a match for every image pixel.

1) *Sparse Methods*: In order to establish sparse stereo correspondences, features must be extracted and then described. We extract and describe Lowe's multiscale SIFT features, as well as Harris corners which are encoded using

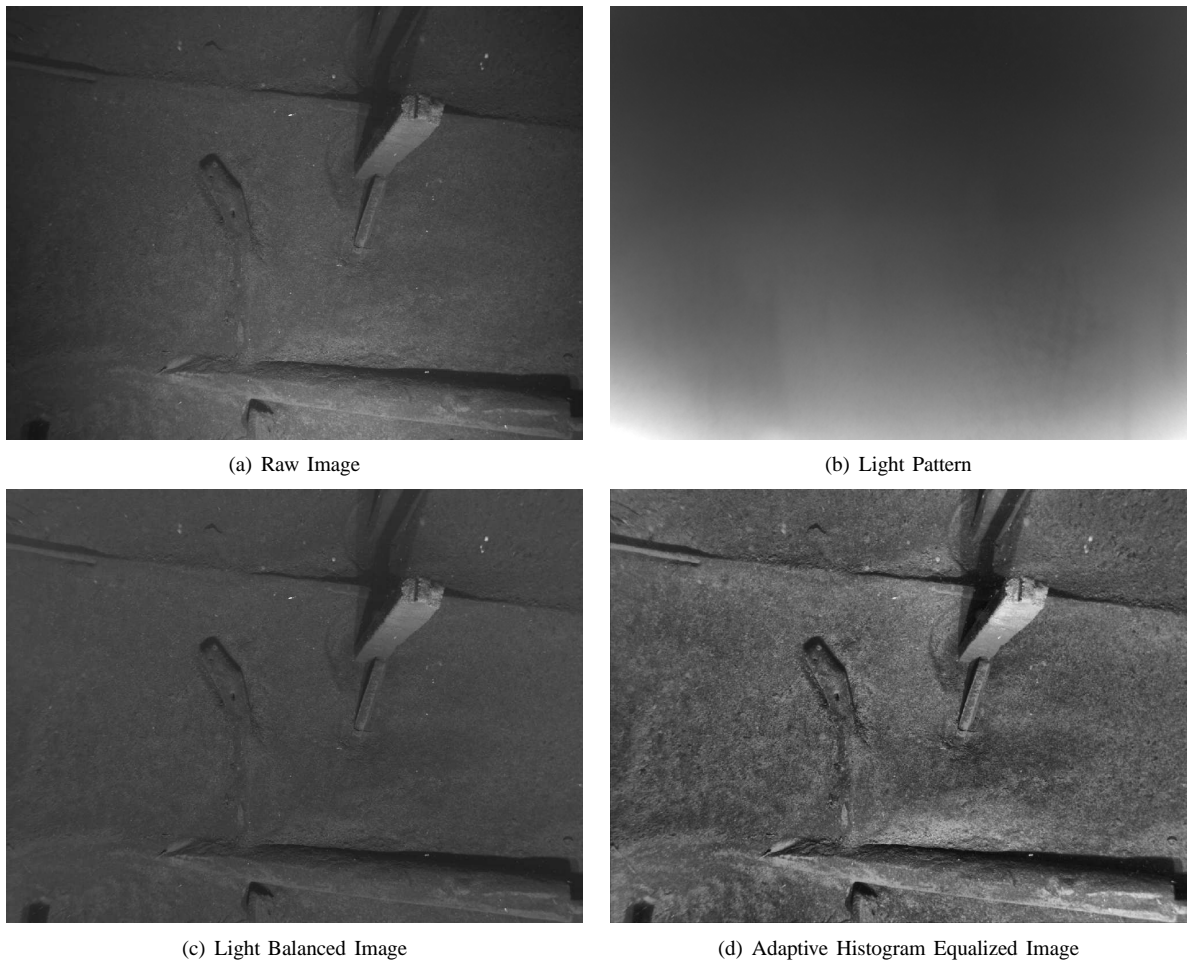


Fig. 3. Image pre-processing steps. A raw image (a) is light balanced using a lighting model (b) to produce image (c). Adaptive histogram equalization accentuates texture (fig. 3(d)).

square windows surrounding the feature [21], [22]. Then a search for inter-image feature matches can be implemented by selecting matches which maximize a similarity measure calculated by comparing feature descriptors.

During a typical exhaustive search each feature in one image is compared to all the features in the opposing image. For calibrated sparse stereo correspondence, the potential correspondences to any one feature in the stereo pair can only lie within a strip around the related epipolar line in the opposing stereo image. Furthermore, if there is some knowledge of scene, the mapping of homogeneous image coordinate u to homogeneous image coordinate u' in the opposing image can be found by using the point transfer function developed by Hartley and Zisserman [8].

$$u' = \frac{H_{\text{inf}}u + K't/Z}{H_{\text{inf}}^{3T}u + t_z/Z} \quad (1)$$

Where H_{inf} is the infinite homography that warps the normalized image coordinates in the first image into the rotational frame of the left image $H_{\text{inf}} = K'RK^{-1}$. The expression H_{inf}^{3T} refers to the third row of the infinite homography and t_z refers to the third element of the translation vector between

the two cameras.

The search region is limited to an error ellipse where the size and orientation are dictated by the uncertainty in the calibration of the stereo rig and the sonar range measurement (fig. 5). This error ellipse is calculated from the propagation of the covariance of the rig calibration and sonar measurements through the point transfer function via the Jacobian matrix of the transformation.

$$C_{u'} = J^T C_{\gamma} J \quad (2)$$

where C_{γ} is the covariance matrix of the stereo calibration and sonar range uncertainty, J is the Jacobian matrix of the point transfer mapping and $C_{u'}$ is the covariance of the image point after it has been mapped onto the opposing image plane. Our knowledge of range, Z in equation 1, comes from the sonar range data which is from a local bathymetry map created with the multibeam sonar. The scene depth at each feature is obtained from the map and backprojected into the camera coordinate system using the camera projection model.

During the correspondence search, a feature extracted from the left hand image is selected, and a corresponding sonar point is selected by the search. The stereo correspondence

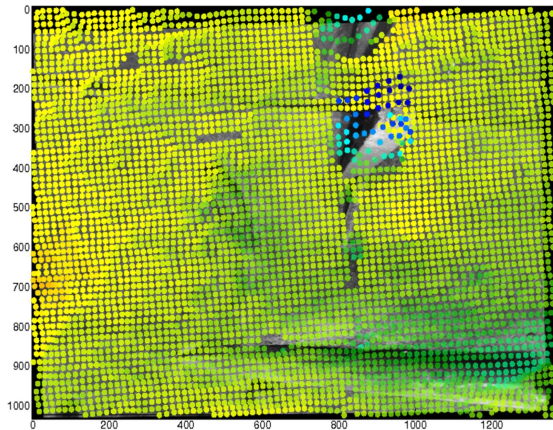


Fig. 4. The backprojected multibeam range data layered on the image for guiding sparse stereo correspondence.

search region in right hand image is selected using the point transfer function and the sonar based estimate of range.

Matches are selected by the feature with the highest similarity score within the search region in the right image. Additionally, a uniqueness constraint is employed wherein the best similarity score is required to be 1.5 times the next best similarity score or the match is eliminated. This is similar to Lowe’s prescribed use of a uniqueness threshold in matching SIFT features[21]. After searching for correspondences to each of the points in the left image, the process is repeated in reverse for features in the right image. Only the potential matches that achieve the maximum similarity score and meet the uniqueness threshold in both directions are retained.

Sift features are costly to extract and describe even though their encryption is a fairly compact 128-vector. They are a very rich and descriptive feature and are often effective in exhaustive searches. We are interested in terrain constrained correspondence as a potential way to reduce the computational demands of extraction and description by allowing for the use of simpler features than SIFT or Zernike Moments, which have been used in previous work with underwater sparse stereo [14]. Therefore we test similarity scores based on normalized cross-correlation which has fewer invariances than SIFT descriptors and is less likely to find correct unique matches in an exhaustive search. Similarity scores were based on cross-correlation of square windows surrounding the Harris corners and correspondence is awarded to the highest similarity score.

These matches can then be triangulated to form a sparse three dimensional reconstruction of the scene, which, for final visualization can be gridded, smoothed and texture mapped with any image taken of the location.

D. Dense methods

Dense correspondence searches try to find stereo matches for each pixel in the image, regardless of photometric interest

level. Such searches are typically simplified by rectifying the images such that corresponding epipolar lines are parallel and appear on the same image scanline, thus reducing the search from a 2D to a 1D search[23]. We propose to constrain the search further by setting a minimum and maximum disparity search limit based on sonar range data.

First, new projection matrices must be calculated which project the image planes onto to new parallel, rectified image planes. In rectified images depth is related to pixel disparity, or difference in horizontal position of two matching pixels in each image, as follows:

$$R = bf/d \quad (3)$$

where R is range, b is the baseline between cameras, f is the focal length, and d is the pixel disparity.

In block-matching stereo algorithms, the scene approximation is manifested as the minimum and maximum disparity search range dictating how many pixels in the right image are compared with each pixel in the left[20]. When more complete knowledge of scene is available, this number can be reduced and the search can be more focused on the region in the right hand image most likely to contain the correspondence.

After image rectification, terrain data must be back-projected onto the left image plane using the rectified projection matrices. There are different ways for utilizing terrain-based estimate of disparity. One possible way of using this information is to tile the estimated disparity map and select the minimum and maximum ranges that fall into each tile to be the minimum and maximum search pixels in each corresponding tile in the left hand image. A z-buffer is another way to apply the sonar ranges to a disparity estimate. It allows for further constraint, especially in the case of considerable depth changes where tiling may require the search of a considerable range of disparities. To create a z-buffer, the range data must be tessellated before back-projection to resolve occlusions. For both of these methodologies, the result is an estimate of disparity keyed to the left image.

Once an initial guess of disparity is made, a terrain derived, normalized cross-correlation block matching algorithm is used. This method is similar to the block matching employed in OpenCV where pixels are matched based on a Sum of Squared Differences similarly measure [23]. However, because we have knowledge of the scene, we can launch a smaller and more directed correspondence search.

One characteristic of dense correspondence is the large number of algorithmic parameters such as disparity search range, and window size as well as textural, uniqueness, and speckle rejection criteria which can be adjusted. We test our method in absence of these user defined inputs which may be added in later iterations on this method. Our intention is to set these by knowledge of terrain as we have already done with minimum and maximum disparity.

III. RESULTS

The sample images used to illustrate the methodologies in (figs. 4, 9, 8, 6) in this paper were chosen because they

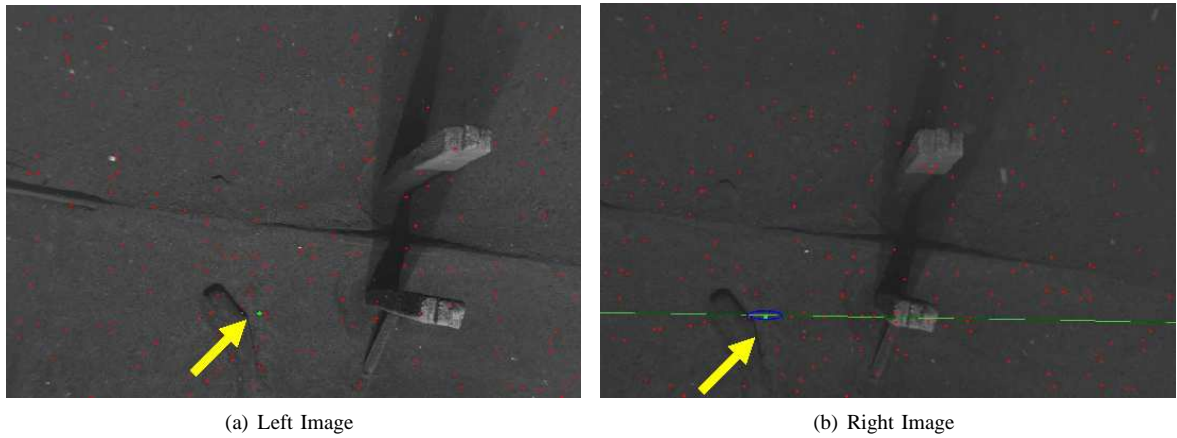
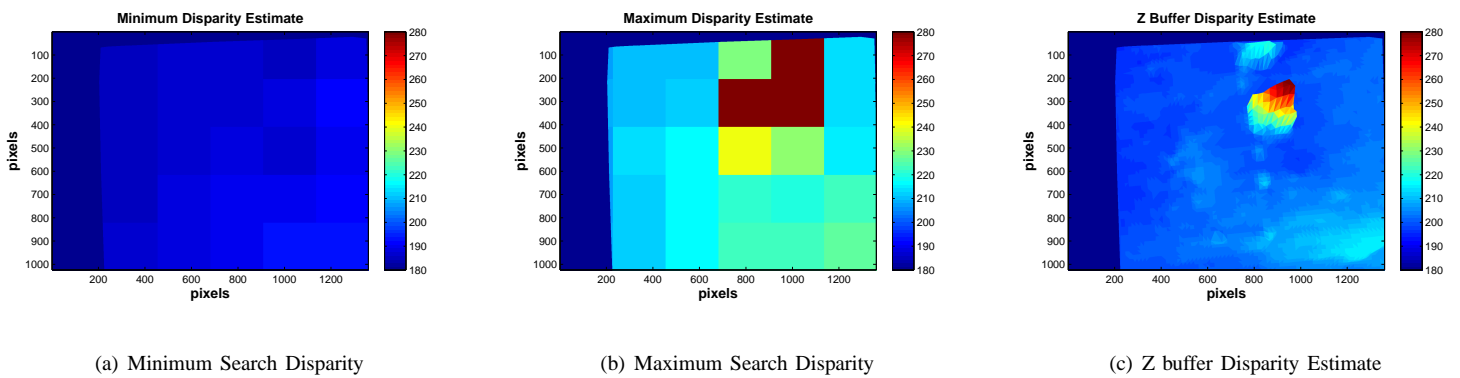


Fig. 5. Sparse terrain constrained correspondence search methodology. Arrows indicate the corresponding features in each image. The ellipse in (b) indicates the search region.



(a) Minimum Search Disparity (b) Maximum Search Disparity (c) Z buffer Disparity Estimate

Fig. 6. Backprojected range data from the scene in (fig. 3) transformed into an estimate of disparity to guide the dense correspondence search.

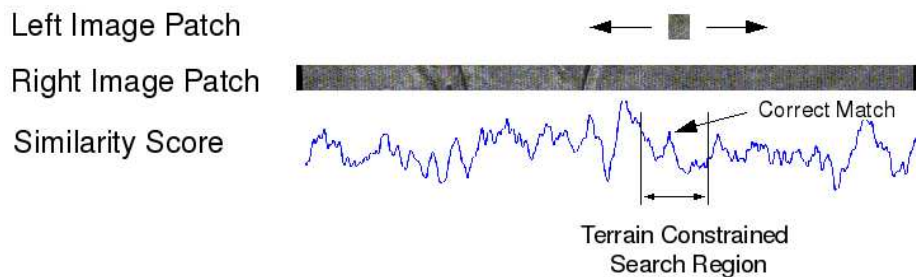


Fig. 7. Dense terrain constrained correspondence methodology. Only right hand image feature patches falling into the constrained search region are compared via normalized cross-correlation to the left image feature patch.

illustrate several characteristics of underwater imagery that pose challenges to stereo correspondence. The terrain has little photometric variation, some back-scatter, and a large depth discontinuity.

A. Sparse Correspondence Results

In the sample stereo pair shown in (fig. 3), 47,425 SIFT features were extracted from the left image and 48,793 features from the right, this covers the image quite densely. An unconstrained search for correspondences using a uniqueness

threshold of 1.5 as recommended by Lowe [21] results in 5,359 matches. Outliers lie both in front of and behind the camera (fig. 8(a)). When the sonar is used to constrain the correspondence search, 8,274 matches are found prior to outlier rejection and smoothing. The outliers lie close to the correct matches and do not obscure the structure of the scene.

When 48,000 Harris corners are extracted and a constrained search is used, 19,135 matches are found. Because the decreased descriptiveness of the feature encryption, made possible due to the usage of terrain constraints, many good

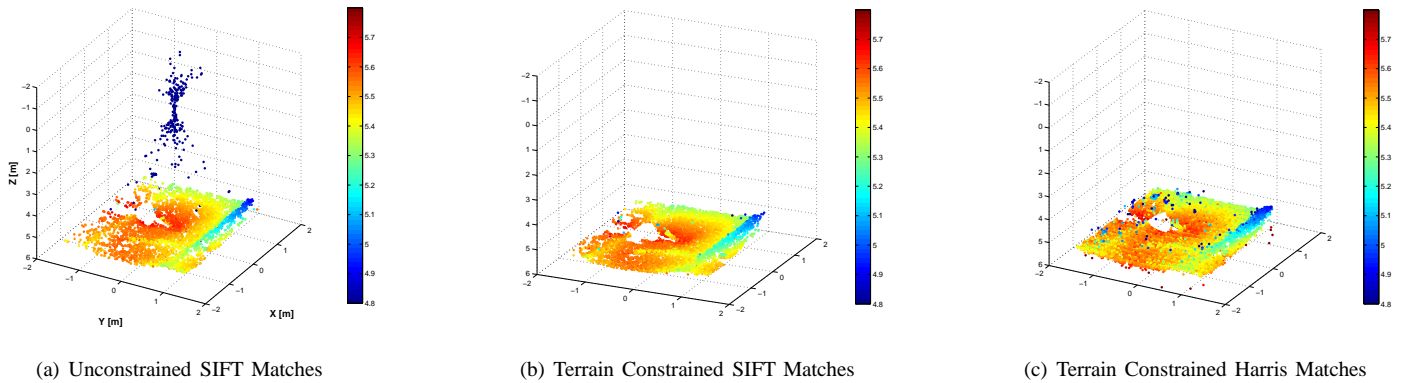


Fig. 8. Results of constrained and unconstrained sparse correspondences. SIFT features were extracted and matched in an exhaustive search (a) and a terrain constrained correspondence search (b). Harris features compared using terrain constrained normalized cross-correlation (c).

matches that may have been rejected with SIFT features are no longer rejected. Thus there are many more matches in this reconstruction. The trade-off is that more bad matches also occur. Note the vertical range ripples in the reconstruction, this is a result of the discretization of the image into pixels along the scanlines. This may be corrected by employing interpolation along the scanline to allow for subpixel matching and resolution [17].

B. Dense Correspondence Results

The first set of dense correspondence trials were done with no initial guess for disparity and a 40×40 search window. Predictably, this is very costly in terms of time since 1360 right-hand image pixels must be compared with each left hand pixel. Additionally, correct matches are difficult to achieve when comparing 1360 similarities calculated from normalized cross correlation. The structure of the scene is readily discernible but the low texture regions contain a large number of mismatches (fig. 9(a)).

Next we set one minimum and one maximum value for estimated disparity. These parameters remain constant across the whole image. This is comparable to the Block Matching implementation used in OpenCV[23]. The minimum disparity selected was 180 pixels for this particular stereo pair and the maximum was 280 pixels. This is a search of 100 pixels and is more likely to contain a unique similarity maximum than the 1360 pixel search. (fig. 7). The resulting dense map is shown (fig. 9(b)). For this test, a 40×40 search window was used. Smaller search windows work best for highly textured scenes whereas larger search windows work best for less textured scenes[17]. Because these images are, relatively speaking, low in texture, we found that a larger window gave the best results. Our chosen window size is similar to, but on the upper end of window sizes found in previous work when image size is taken into account[12]. The resulting dense map is shown (fig. 9(b)). False matches are more prevalent in areas with less textural information and speckle is apparent in the areas around occlusions and large depth discontinuities.

To further constrain the search, we set the minimum and

maximum search disparity based on tiled minimum and maximum disparity estimates from the backprojected sonar (fig. 6). This resulted in a faster correspondence search since no more than 100 pixels were searched in the opposing image and often as few as 10. In some areas of low texture where mismatches occurred in the first two trials, they reoccur in the constrained searches, but these pixels are constrained in such a way that scene structure isn't obscured (fig. 9(c)). In other regions, the search has been constrained such that correct matches are now found where mismatches occurred previously.

To further constrain the search using terrain, the bathymetry is back-projected onto the image plane using a z-buffer algorithm to resolve occlusions. This is the most computationally expensive method of estimating disparity (fig. 9(d)). We then search 20 pixels on either side of the estimated disparity. The z-buffer is the most restrictive constraint we can impose on the correspondence search, however, it is more reliant on the accuracy of the coregistration between the two data sets. The figure shows increasing ability to cope with outliers in a manner that preserves scene geometry. Additionally there are regions where number of correct matches are increased with increasing constraint on disparity search region.

C. Back-Projection of Terrain

When the bathymetry is back-projected onto the image plane, we find a certain amount of misalignment between the sonar data set and the image. This is a result of inadequate knowledge of the rigid transformation between the sonar and the cameras. The search region corresponds to a random position variable with a chi-squared, two degree of freedom distribution. However this imperfect registration is a deterministic offset present everywhere in the back-projected range which can't be accommodated by the point transfer mapping. It may eliminate correct matches by casting them out of the search region and isolating incorrect matches.

An advantage of having coregistered bathymetry and stereo is that we can compare the two data sets. This will help us gain a better understanding of the relative strengths and limitations of the two sensing modalities (fig. 10). The figure shows the

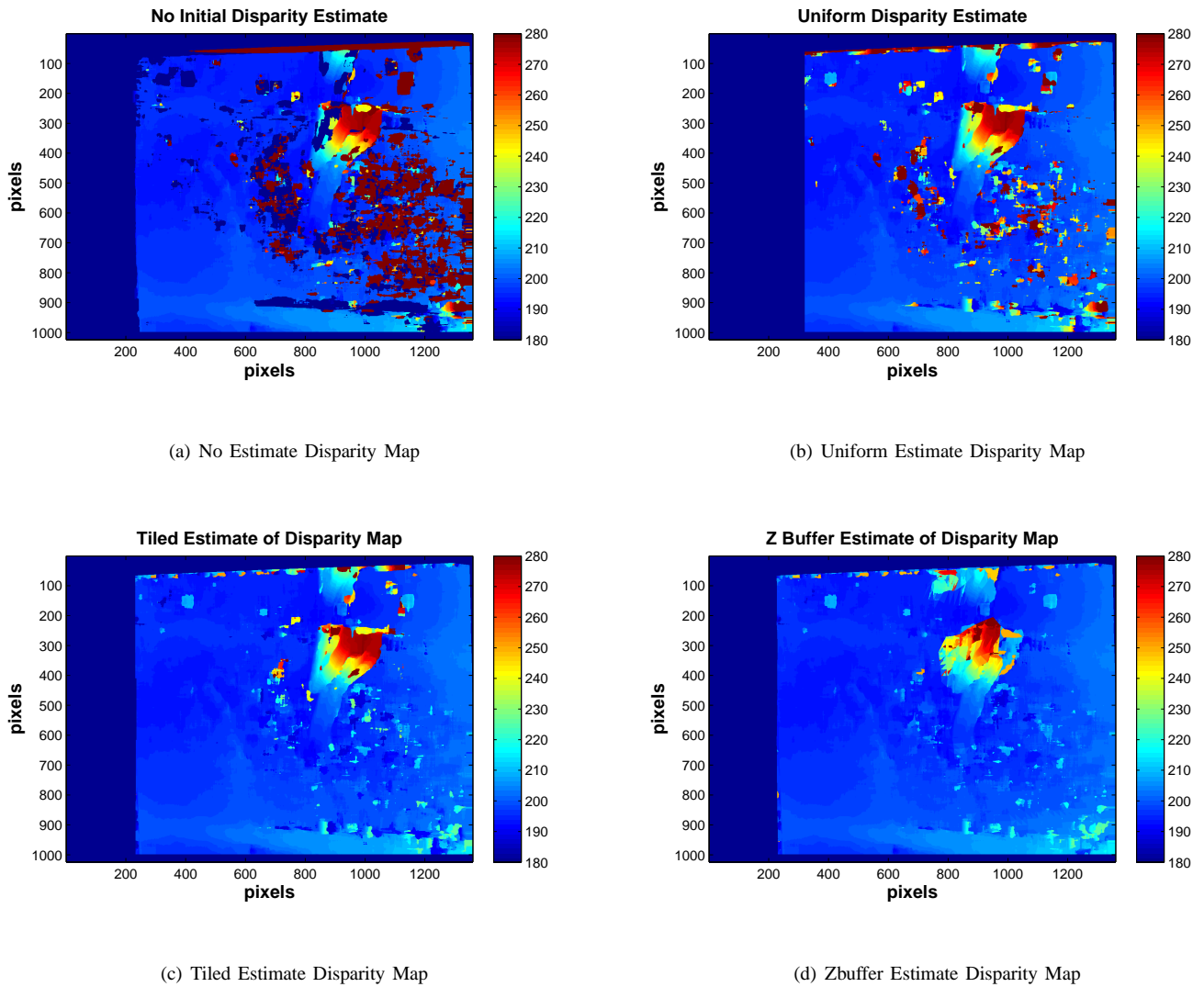


Fig. 9. Disparity maps based on several different initial disparity estimates from multibeam sonar terrain model.

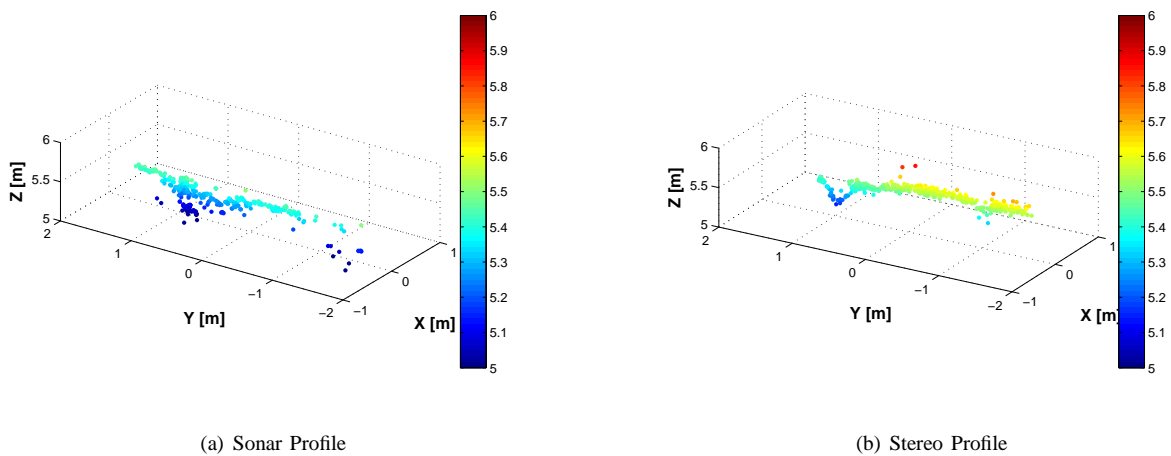


Fig. 10. Two range thin-sections from Sinope D. The coregistration of the sonar and stereo data sets allows for comparison between the two modalities.

same thin section as represented by each sensor.

IV. CONCLUSION

We have developed a method for fusing sonar and stereo imaging sensing modalities for creating three dimensional sea floor maps. Local micro-bathymetry is associated with corresponding image points through back-projection onto the image plane where it is used to constrain the search region for a given feature's match in the opposing image. This methodology can be applied in both sparse and dense correspondence search frameworks to produce depth maps.

In the case of sparse matching, we are able to match a large number of features to gain a high density point cloud of the sea floor wherein many outliers are rejected by the terrain constraint. The structure of the sea-floor is apparent in spite of outliers.

In low contrast images dense correspondence search is improved by increasing the level of terrain constraint. In some cases correct matches are selected where unconstrained searches failed, and in some cases, incorrect matches continue to be selected, but are constrained to prior knowledge of terrain in such a way that they don't obscure scene structure in the resulting depth map.

Future work will address the limitations caused by the deterministic mis-registration between the sonar and the cameras, which results from limited knowledge of the relative sensor transforms.

A product of the coregistration of acoustic and optical bathymetry is the the sonar data can be directly substituted in areas where stereo fails and vice versa. It is sometimes difficult to discern where the matches have failed, except by observation, therefore it maybe advantageous to incorporate textural and speckle rejection constraints as criteria for the adoption of sonar range data in place of stereo. It also may be helpful in the future to use the bathymetry as a principled way of sidestepping the heuristic portion of parameterizing the dense stereo methodology, by terrain knowlege to set window size or another parameter.

The juxtaposition of the two datasets also provides an opportunity to quantify the characteristics of the two sensing modalities in order to better understand and exploit their relative merits.

ACKNOWLEDGMENT

The authors would like to acknowledge The Institute for Exploration and the Archeological Oceanography 2007 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," 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, Sept. 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. Papadopoulos, *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 January, 2008. [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc/index.html

[17] K. Won, A. Ansar, R. Steele, and R. Steinke, "Performance analysis and validation of a stereo vision system," in *IEEE International Conference on Systems, Man and Cybernetics*, October 2005.

[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] H. Singh, O. Pizarro, L. Whitcomb, and D. Yoerger, "In-situ attitude calibration for high resolution bathymetric surveys with underwater robotic vehicles," *Robotics and Automation, 2000. Proceedings. ICRA '00. IEEE International Conference on*, vol. 2, pp. 1767–1774, 2000.

[20] K. K. and D. Beyer, *SRI Small Vision System: User's Manual*. SRI International, May 1997.

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

[22] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of the Fourth Alvey Vision Conference*, vol. 1, 1988, pp. 147–151.

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