Underwater Object Reconstruction Method for Target Recognition in an Autonomous Underwater Vehicle

Pavan Kumar Ambekar#, Mutanna Kadal&, Karthik Sheshadri*
#* Dept of Computer Science, RVCE, India
&Dept of Aerospace Engineering, IISc, India
e-mail: #pavan.gma950@gmail.com, &muttanna82@gmail.com, *karthiksheshadri@gmail.com

***Abstract-* We consider the problem of reconstructing 3D models of underwater objects from multiview images and propose a low complexity algorithm to achieve this without loss in accuracy as compared to existing work. We utilize bicubic interpolation as a substitute for Gaussian decomposition to arrive at multiresolution images. We use Kolmogorov modeling to generate disparity maps from various views and hence obtain surface terrains which are combined to construct a complete 3D model. We demonstrate that the proposed algorithm entails significant time saving while retaining the structural integrity necessary for the recognition system in an Autonomous Underwater Vehicle (AUV).**

I. INTRODUCTION

The holy grail of research in Pattern Recognition and Artificial Intelligence is the design and construction of computers with perception and awareness comparable to those of human beings. A challenging area of current interest in Computer Vision is the reconstruction of 3D models from multiple 2D images. While extensive research has been carried out in the area of 3D reconstruction [1] – [4], there has not been much work which deals explicitly with underwater objects in general and the stringent real time constraints imposed by an AUV in particular. In this paper we present a framework for rapid 3D reconstruction of underwater objects. Our approach is designed to minimize algorithmic complexity, hence enabling real time application in the object recognition system of an AUV.

Underwater image processing presents various difficulties such as visibility, light absorption, non-uniform illumination, etc [5]. References [5]-[8] investigate several preprocessing approaches to deal with this problem. We utilize the approach presented in [5] since it is the only work that presents a comprehensive treatment of the preprocessing problem. Other approaches address specific issues but fail to provide compatible solutions to the remaining problems.

Our work relates most to the reconstruction methodology presented in [9]. However, this method is time consuming and hence may not be feasible for a real time application like ours. Also, [9] does not present a complete reconstruction of the scene.

We capture images from different angles using a calibrated camera setup [10]. As mentioned earlier, we perform preprocessing as in [5] on the acquired images. The images are rectified based on the intrinsic and extrinsic parameters of the setup [10]. A set of multiresolution images are obtained by bicubic resizing [11]. We obtain disparity maps using the method presented in [12]. The terrain of each view is obtained by surfing based on the intensity of the pixels in the disparity map. The disparity maps obtained for different views are used to generate a complete model using vector addition on the overlapped areas in the disparity maps. We present time comparisons with the approach discussed in [9], showing that the proposed method entails considerable time saving without loss in accuracy.

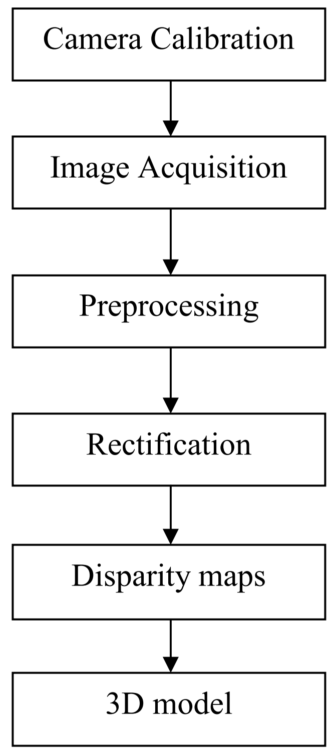


Figure 1. Illustrating the reconstruction process

II. EPIPOLAR RECTIFICATION

As mentioned earlier, underwater images suffer from various problems such as limited range, non uniform lighting, low contrast, blur, loss of colour, etc [5]. The use of raw images would lead to non uniform intensities in the multiresolution images and hence cause inaccuracies in the 3D model. Image enhancement and preprocessing is carried out as in [5] to ensure uniformity of multiresolution images.

The images are acquired from different angles using either multiple cameras or a single camera which can be rotated by

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

preset angles [10]. For theoretical purposes we consider the case of two cameras. The analysis can easily be extended to the case of multiple cameras. When one considers two cameras photographing a common object, one encounters the problem of finding a corresponding point viewed by one camera in the image of the other camera. Image rectification is a process whereby the images from the two cameras are transformed such that their epipolar lines [10] align horizontally. The geometry of the setup is defined by:

(1) the distance between object and camera.
(2) the pixel to mm ratio for each camera angle (k_u, k_v).
(3) the focal length of the camera lens (f).

We utilize the rectification procedure presented in [10]. Accordingly, we define a world co-ordinate system (WCS) where a point w_i is defined by

$$w_i = [x_i, y_i, z_i]^t$$

and a camera co-ordinate system (CCS) where a point m_i is defined by

$$m_i = [u_i, v_i]^t$$

A perspective projection is used to map 3D co-ordinates to 2D co-ordinates as shown below:
Let $\widetilde{m}_i = \widetilde{p}\,\widetilde{w}_i$
where

$$\widetilde{w}_i = [x_i, y_i, z_i, 1]^t$$

$$\widetilde{m}_i = [u_i, v_i, 1]^t$$

are the homogenous co-ordinates of the WCS and CCS respectively, and $\widetilde{p}$ is the perspective projection matrix.
Then we define

$$A = \begin{bmatrix} a_u & \gamma & u_0 \\ 0 & a_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\widetilde{p} = [R|t]$$

where $a_u = -f\,k_u$
$a_v = -f\,k_v$
(u_0,v_0) is the intersection of the optical axis and the retinal plane (principal point).
γ is the skew factor and R and t are the rotation matrix and translation vector respectively.

In [10] it is shown that the optical ray of an image point m, i.e, mc, obeys the Cartesian equation

$$w = c + \lambda Q^{-1}\widetilde{m} \qquad (1)$$

where
c are the coordinates of optical center of the camera
$Q^{-1} = (AR)^{-1}$ and λ is a real number.

This transformation determines an equivalent WCS co-ordinate for every CCS co-ordinate from the stereo images.

Figs. 2 and 3 show the geometry of the setup prior to and after rectification respectively when two cameras are used to capture images of a point w. E_i is the epipole of the i'th view and M_i is the point on the i'th image plane corresponding to w, and C_i denotes the optical centre of the i'th camera.

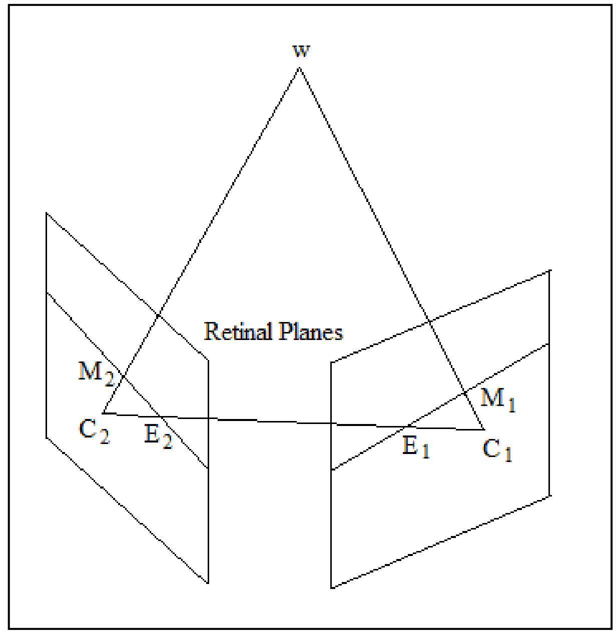


Figure 2. The case of unrectified camera angles

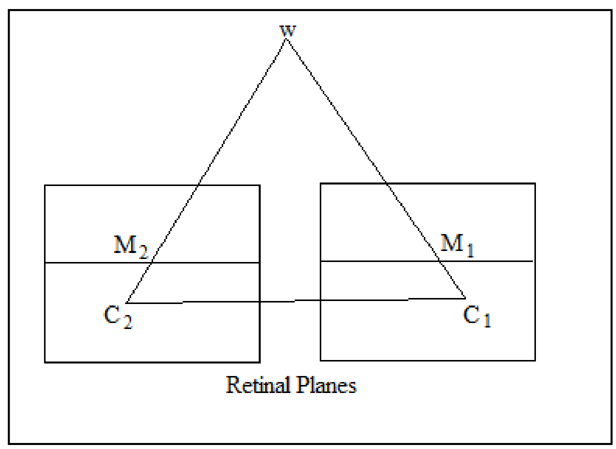


Figure 3. The rectified image planes of two cameras

III. Obtaining Multi-resolution Images

In this work we use bicubic interpolation as a means of procuring multi resolution images from the rectified images. Bicubic interpolation is an extension of cubic interpolation for interpolating data points on a 2D grid [11]. Images resampled with bicubic interpolation are smoother than those obtained by bilinear interpolation or the nearest neighbor approach and have fewer interpolation artifacts.

In our work we utilize the bicubic convolution algorithm [11]. Accordingly, the resultant image is

$$z(x) = \begin{cases} (a+2)\,|x|^3 - (a+3)\,|x|^2 + 1, & |x| \le 1 \\ a|x|^3 - 5a|x|^2 + 8a|x| - 4a, & 1 < |x| < 2 \\ 0, & \text{otherwise} \end{cases} \qquad (2)$$

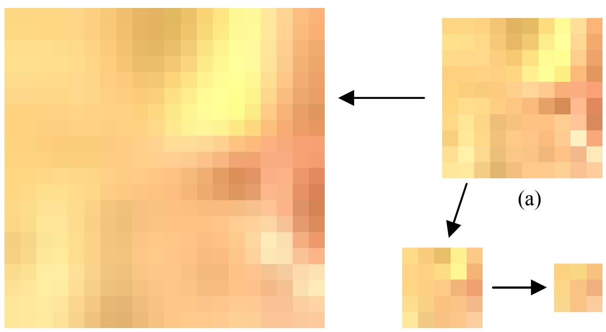
where x is a spatial pixel

z(x) denotes the set of pixels obtained from applying the bicubic interpolation process on the pixel x and

a is a constant parameter which we set to –0.5 for the purposes of this work.

An alternative approach is to use a Gaussian pyramid [13]. We have found that the use of bicubic resizing yields similar results with considerable time saving.

Fig. 4 shows bicubic interpolation on a 10X10 sized pixel patch 4(a) to produce 20 X 20 sized 4(b), 5 X 5 sized 4(c) and 3 X 3 sized 4(d).

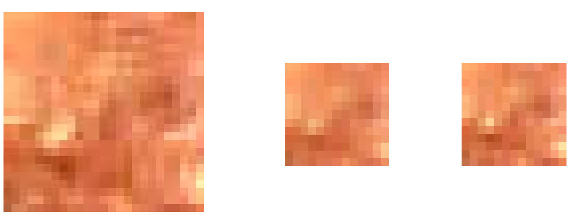


(a)

(b) (c) (d)

Figure 4. Bicubic interpolation performed on a 10X10 segment (a) of an image of a boulder

Fig. 5 shows bicubic and Gaussian interpolation on a 25X25 pixel patch of the boulder 5(a). The time taken for producing a 13X13 resized image by the Gaussian pyramid method (5(b)) was 0.022559 s and the time taken for producing the same by bicubic interpolation (5(c)) was 0.010196 s.



(a) (b) (c)

Figure 5. Gaussian (b) and bicubic (c) interpolation performed on a 25X25 segment (a) of an image of a boulder

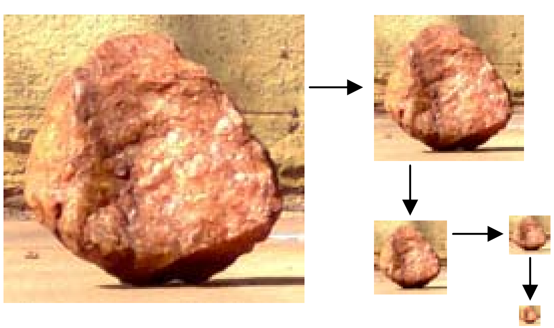


Figure 6. The bicubic interpolation sequence shown on an image of a boulder

We carry out stereo matching by the sum of squared differences method [4] on each multiresolution image obtained from the bicubic interpolation process.

IV. Generation of Disparity Maps

In sections II and III we have shown that we may obtain several multiresolution images of the rectified images using a calibrated camera system. Each camera sees a slightly different projection of the world view. We may compare these images to obtain relative depth information of points in the object, which is referred to as a disparity map.

We obtain the disparity map using the method proposed in [12], which uses an energy minimization technique as explained below:

Let P_i be the set of pixels of the image obtained from the i'th camera. We define

$$P = \bigcup_{i=1}^{n} P_i \quad (3)$$

Since the objective is to estimate depth in the image, we need to obtain the point of "first intersection" between a pixel P_i and any object in the scene.

In other words, we have to obtain a labeling f: P ⟶ L where L is a discrete set of labels corresponding to different depths.

A point p in 3D space can be represented as a pair $< p,l >$ where $p \in P$ is a pixel at a depth $l \in L$.

Let a set I consist of unordered pairs of 3D points $<p_1,l_1>$ and $<p_2,l_2>$ with the points being in same depth $l_1 = l_2$

Reference [12] considers an energy function which addresses symmetry, visibility and smoothness as the various disparate issues to be addressed, quantising each as in the equation below:

$$E(f) = E_{data}(f) + E_{smoothness}(f) + E_{visibility}(f) \quad (4)$$

where

$$E_{data}(f) = \sum_{<p,f(p)>,<q,f(q)> \in I} D(p,q) \quad (5)$$

where $D(p,q) = \min\{0, (Intensity(p), Intensity(q))^2 - K\}$

p and q are pixels in I

$K > 0$ is a constant

$$E_{smoothness}(f) = \sum_{\{p,q\} \in N} V_{\{p,q\}}(f(p), f(q)) \quad (6)$$

where N <subset> ⊂ {{p,q} : p,q ∈ P}

$V(l_1,l_2) = \min(|l_1 - l_2|, K`)$ for constant K`

We define the set I_{vis} with the following condition if $\{<p_1,l_1>,<p_2,l_2>\} \in I_{vis}$ then $l_1 \neq l_2$ ie. points at different depths. If the following constraint is satisfied $E_{visibility}(f)$ is 0, otherwise it is ∞.

$$E_{visibility}(f) = \sum_{<p,f(p)>,<q,f(q)> \in I_{vis}} \infty \quad (7)$$

Reference [12] gives an approximation algorithm which we use to minimize the energy function.
The term "graph cut" is explained below:
Let G = <V,E> be a weighted graph with terminal vertices {s,t} called source and sink respectively. A cut C = V^s, V^t is a partition of the vertices into two sets such that $s \in V^s$ and $t \in V^t$.
The cost of the cut, $|C| = \sum w(e)$
where w(e) is the weight of the edge(e) between a vertex V^s and a vertex in V^t.

V. OBJECT RECONSTRUCTION

The models are obtained from the disparity maps constructed using the above method. The fact that the intensity gradient in a disparity map describes the contour of the object in a particular view can be used to generate the terrain of that particular view. This can be done by surfing over the surface of the disparity map proportional to the depth associated with each pixel as shown in the models Fig. 7(d), Fig. 9(b), Fig. 12, Fig. 13(d) and Fig. 14(a).

A. Generation of the complete 3D model

The problem of arriving at a complete 3D model can be addressed as below: In section IV we have shown that disparity maps and hence terrains may be obtained from each view.

One method of obtaining a complete model is to simply overlap the disparity maps. However the region of overlap seen from the view directly in front is not constructed accurately (Fig. 7). Instead, we generate this region using vector addition of the two overlapping views as illustrated in Fig. 8.

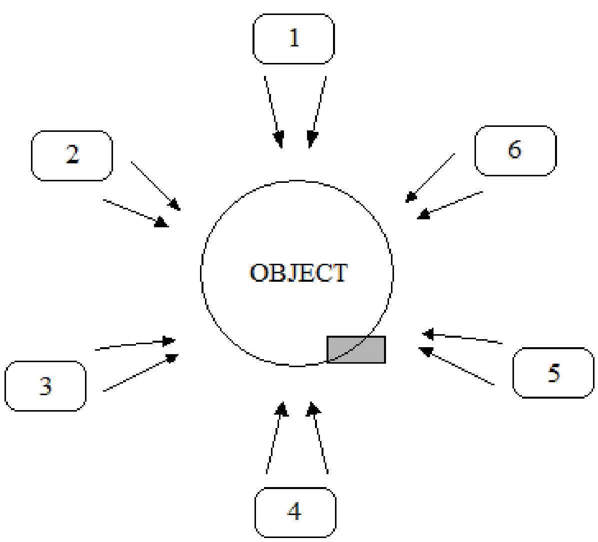


Figure 7. Illustrating the problem of overlapped regions.

The characteristics of the terrain depend on two factors:

1. The levels, l_i and l_j for point p in the overlapped region in the disparity maps of the n^{th} and $(n+1)^{th}$ views respectively.
2. The spatial coordinates of the pixel in the disparity maps of both the images.

Both the above factors can be represented by a vector for each overlapped pixel. The magnitude and direction of this vector are given by the level and position of the corresponding pixel in the disparity map respectively, as shown in Fig. 8. We thus obtain an approximate terrain of the overlapped area.

Reducing the number of disparity maps required for a complete view significantly reduces time required.

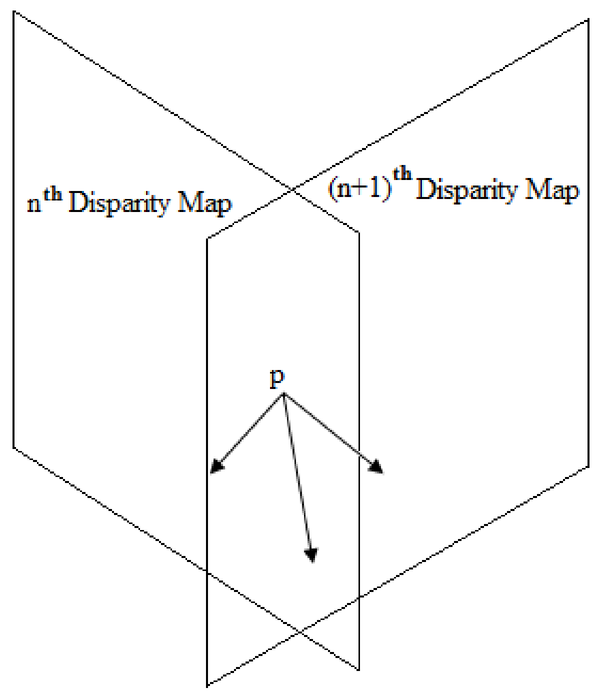
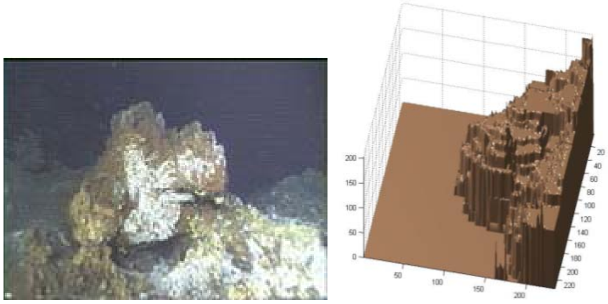


Figure 8. Pixel p is reproduced in the new approximate disparity map in the overlapped region by vector addition as explained.

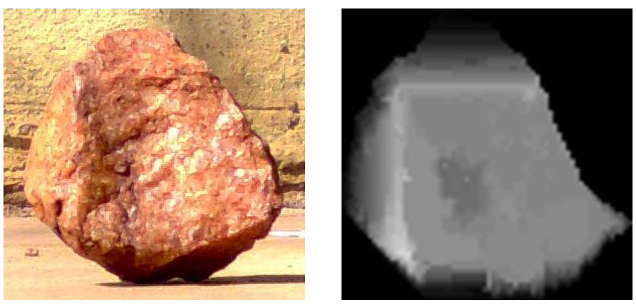
VI. SIMULATION AND RESULTS

In this section we present results from a simulation of the proposed reconstruction method on various stereo images. We carry out the proposed procedure on both underwater images and on stereo images of a boulder. We also present time comparisons with existing procedures (Gaussian resizing and Delaunay triangulation). Our simulation was carried out on a 2.0 GHz Intel Core2duo processor T6400 with 3GB DDR3 RAM and an Intel Graphic Media Accelerator 4500MHD. The tables below compare the computing times for each model using Delaunay triangulation and surfing.

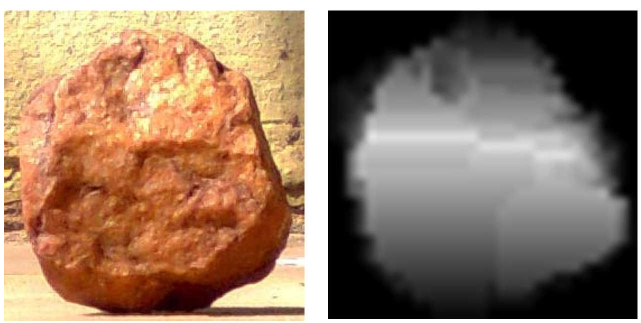


(a) (b)

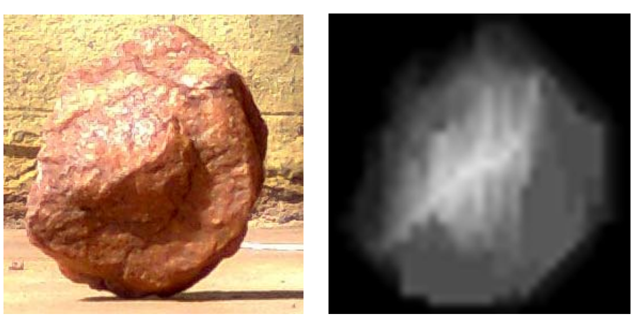
Figure 9. Object reconstruction from the stereo images in [9]



(a) (b)

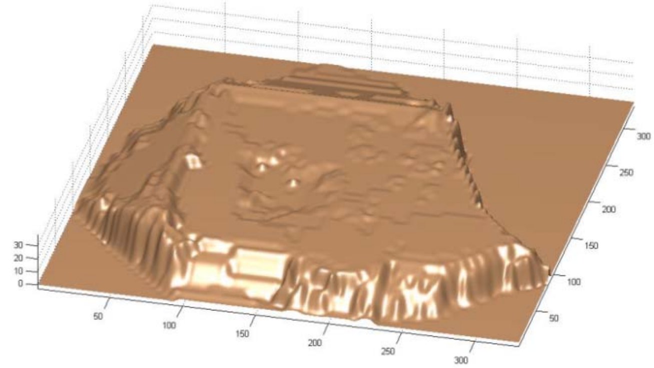


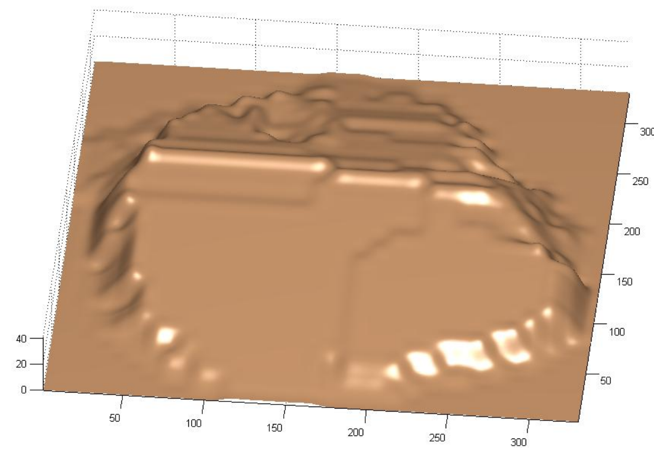
(c) (d)



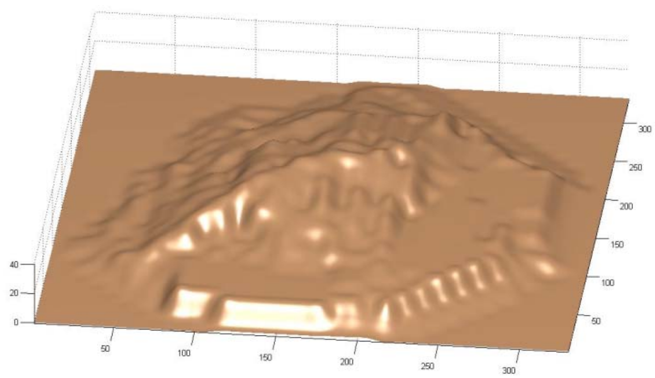
(e) (f)

Figure 10. Disparity Maps (b), (d) and (f) of different views of the boulder (a), (c) and (e) respectively



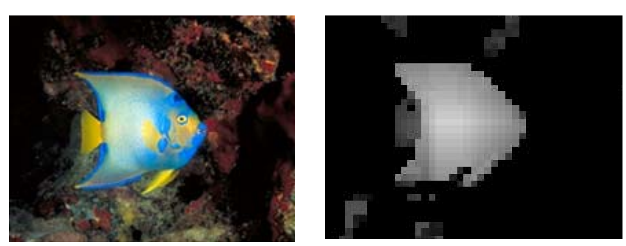


(b)

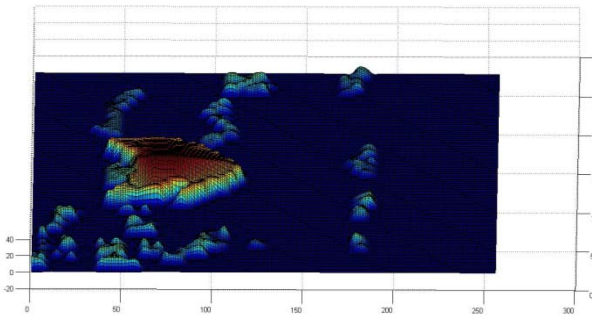


(c)

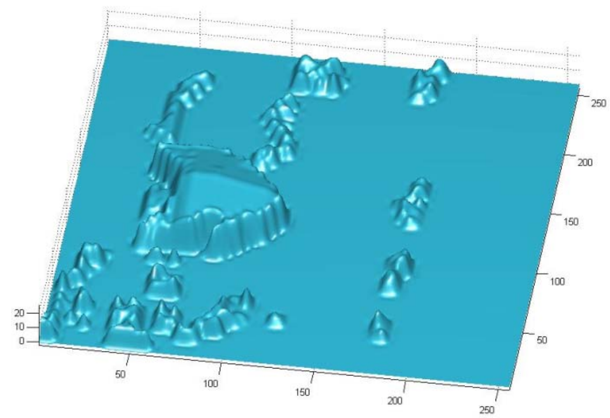
Figure 11. 3D terrains (a), (b) and (c) corresponding to Figures 11(a), 11(c) and 11(e)



(a) (b)



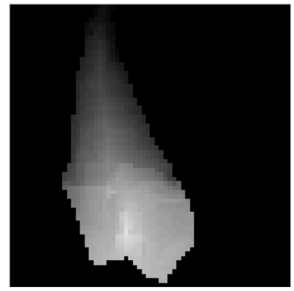
(c)

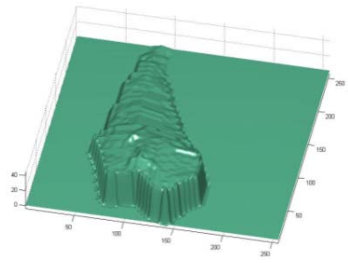


(d)

Figure 12. Object reconstruction from binocular images of a fish [14]

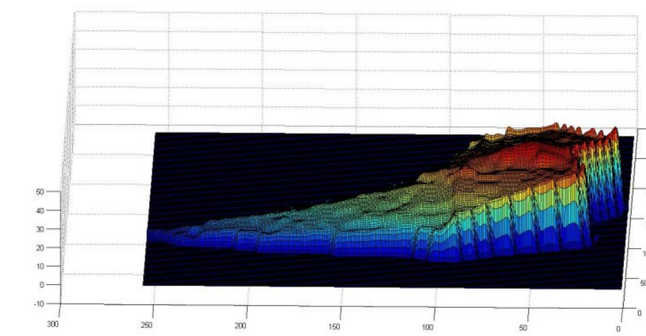
(a) (b)



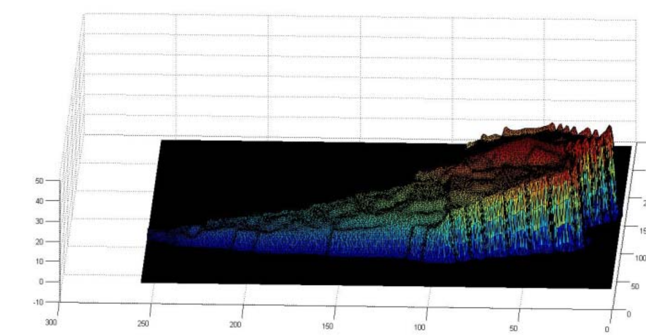


(c) (d)

Figure 13. (a) and (b) are the binocular images of a sea creature [14]. (c) is its disparity map and (d) is the 3D model.



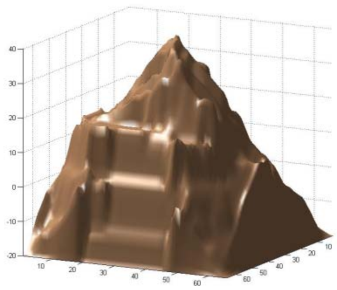
(a)



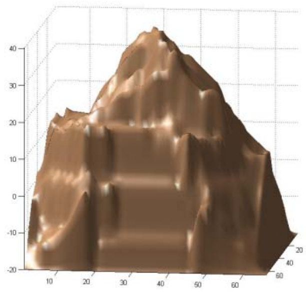
(b)

Figure 14. Visual models of Fig. 13(a) and 13(b) generated by trivial surf (a) and delaunay triangulation algorithm (b)

(a)



(b)



(c)

Figure 15. The complete 3d model of (a) is shown in (b) and (c)

TABLE I
TIME COMPARISON OF GAUSSIAN AND BICUBIC INTERPOLATION PERFORMED TO REDUCE THE RESOLUTION TO HALF THAT OF THE ORIGINAL IMAGE

Fig. used as input	Resolution	Time taken for Gaussian pyramid (in s)	Time taken for Bicubic interpolation (in s)
5(a)	25 X 25	0.022559	0.010196
9	147 X 195	0.161077	0.022777
10(a)	298 X 300	0.459760	0.066937
10(c)	277 X 268	0.404277	0.044362
10(e)	279 X 281	0.413491	0.047197
12	165 X 208	0.193593	0.025793
13	316 X 212	0.364869	0.050868

TABLE II
TIME COMPARISON OF SURFING AND DELAUNAY ALGORITHMS

Input Fig.	Resolution	Time taken for Delaunay triangulation (in s)	Time taken for Trivial Surf (in s)
9	234 X 234	2.541561	0.049449
10(b)	330 X 330	5.262402	0.072819
10(d)	330 X 330	5.262160	0.054908
10(f)	330 X 330	5.184842	0.079447
12	256 X 256	3.660804	0.064049
13	256 X 256	3.056820	0.047034

VII. CONCLUSION

In this paper we review the state of the art in underwater object modeling and address the problem of modifying these approaches to suit the real time constraints of an AUV. We present results that demonstrate drastic time saving over existing approaches without perceptible loss of accuracy.

REFERENCES

[1] Sadjadi, F. ,"Passive 3D Reconstruction using Polarimetric Infrared Imagery",Lasers and Electro-Optics Society (LEOS) 2007, October 2007.

[2] Gurram, P. Rhody, H. Kerekes, J. Lach, S. Saber, E., "3D Scene Reconstruction through a Fusion of Passive Video and Lidar Imagery", Proceedings of the 36th IEEE Applied Imagery and Pattern Recognition Workshop, March 2008.

[3] Lhuillier, M., Quan, L., "Surface Reconstruction by Integrating 3D and 2D Data of Multiple Views", 9th IEEE International Conference on Computer Vision, October 2003.

[4] Park, S., Subbarao, M., "A multiview 3D modeling system based on stereo vision techniques", Machine Vision and Applications, Volume 16 pp 148-156.

[5] Bazeille, S., Quidu, I., Jaulin, L., Malkasse, J., "Automatic Underwater Image Pre-Processing",Caracterisation Du Milieu Marin, October 2006.

[6] Arnold-Bos A., Malkasse J., Kervern G., "A Preprocessing Framework for Automatic Underwater Image Denoising", Proceedings of The European Conference on Propagation and Systems, Brest, France, March 2005.

[7] Schechner Y., Karpel N., "Clear Underwater Vision", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2004, 1:536-543, 2004.

[8] Jaffe J., "Underwater Optical Imaging: Status and Prospects", Scripps Institution of Oceanography, 14(3): 66 – 76, 2001.

[9] Brandou, V, Allais, A. G., Perrier, M., Malis, , E., P. Rives, Sarrazin, J., P. M. Sarradin, "3D Reconstruction of Natural Underwater Scenes Using the Stereovision System IRIS", IEEE OCEANS 2007.

[10] Fussiello A, Trucco E, Verri A (2000) A compact algorithm for rectification of stereo pairs. Machine Vision and Applications 12:16–22.

[11] http://en.wikipedia.org/wiki/Bicubic_interpolation

[12] Kolmogorov, V, Zabih, R, and Gortler, S. "Generalized Multi-Camera scene Reconstruction Using Graph Cuts", Fourth International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition, July 2003.

[13] Burt P., Adelson E., "The Laplacian pyramid as a compact image code.", IEEE Transactions on Communications 31(4): 532–540.

[14] http://www.undersea3d.com/

[15] Roy, S. and Cox, I. J. "A maximun-flow formulation of the n-camera stereo correspondance problem", Proceedings of 6th International Conference on Computer Vision, pages 492 – 499, 1998.

[16] J. Bloomenthal, An Implicit Surface Polygonizer, Graphics Gems IV, pp.324-349.

[17] R. Bergevin, M. Soucy, H. Gagnon, and D. Laurendeau, "Toward a general multi-view registration technique," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18, No. 5, May. 1996.

[18] Sarrazin, J, Sarradin, P-M, and the Momareto cruise participants (2006). "Momareto: a cruise dedicated to the spatio-temporal dynamics and the adaptations of hydrothermal vent fauna on the Mid-Atlantic Ridge", *InterRidge News*, Vol 15, pp 24-33.

[19] http://www.dpreview.com/learn/?/key=interpolation