An Expectation-Maximization Framework for the Estimation of Bathymetry from Side-scan Sonar Images

E. Coiras*, Y. Petillot, D.M. Lane, Ocean Systems Laboratory

Heriot-Watt University
Riccarton
Edinburgh, EH14 4AS, UK
*e.coiras@hw.ac.uk

***Abstract* - In this paper a new procedure for the computation of seabed altitude information from side-scan sonar data is presented. Although side-scan sensors do not provide direct measures of seabed elevation, their images are directly related to seabed topography. Using a mathematical model for the sonar ensonification process, approximations to the seabed characteristics can be inferred from the sonar image. The problem is however severely under-constrained, in the sense that not all the parameters involved in the image formation process can be directly determined from the side-scan image. To overcome this difficulty, we propose the utilization of a multi-resolution expectation-maximization framework to select the most probable parameters from the solution space. At every iteration the estimated solution is used to simulate a side-scan image of the observed scene, which is then be compared to the side-scan image actually observed; solution parameters are then refined using gradient-descent optimization. The process is repeated until convergence is achieved.**

I. INTRODUCTION

Side-scan sonar is one of the most widely used imaging systems in the underwater environment. It is relatively cheap and easy to deploy, in comparison with more powerful sensors like multi-beam systems or synthetic aperture sonar. Unlike them however it is unable to directly gather seafloor depth information. The possibility of overcoming this limitation is therefore of great interest for the marine community.

In this paper we describe a multi-resolution expectation-maximization procedure for the estimation of bathymetry from side-scan sonar images. Initially the sonar ensonification process is approximated by a Lambertian diffuse scattering model, which represents the observed image intensities as a function of the parameters describing the seafloor surface. An expectation-maximization routine [1] iteratively optimizes the estimated parameters until convergence is achieved. The utilization of a multi-resolution approach permits to recover features of a larger scale and improve global convergence.

II. BACKGROUND

There has been limited research in the area of 3D reconstruction from side-scan sonar images [2-7]. The main reasons being the complexity of the full mathematical image formation model and the level of pre-processing required. And in most cases where acquisition of seabed topography is important, attention is driven to more straightforward solutions such as multi-beam bathymetric systems.

Most existing work on seabed reconstruction from side-scan has been mainly qualitative and oriented to obstacle-avoidance for underwater vehicles [2, 4-6]. Other works focus on seabed texture classification or object recognition [7-9]. In all these situations precise descriptions of the seabed topography are not critical.

In general, the fundamental idea behind most reconstruction methods is to determine a model for the ensonification process that is simple enough for the image formation problem to be inverted, obtaining an approximation to the surface gradients, which can be globally described as shape-from-shading methods [3]. Our goal is to extend these methods using a statistical approach to determine the most probable configuration of the seabed topography compatible with the side-scan image actually observed. To this end we use an expectation-maximization framework [10] in order to iteratively refine the set of parameters defining our model.

III. IMAGE FORMATION

A. *Side-scan Sonar*

The side-scan image formation process is briefly sketched in Fig.1. The sensor's acoustic source at *o* produces an ensonification pulse that illuminates the seafloor. Some of the acoustic energy reaching any seabed point $\vec{p}$ is scattered back and can be measured by the sensor. The intensity of the corresponding pixel of the side-scan image will depend on the amount of energy scattered back from the surface point. The pulse is not isotropic, but follows a particular beam-profile Φ that depends on the grazing angle α subtended by $\vec{p}$. The amount of energy scattered back also depends on the seabed reflectivity *R* at the point.

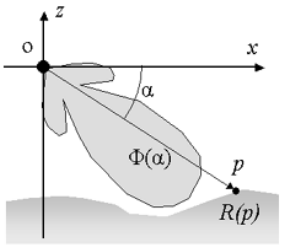

Fig.1: Side-scan image formation.

B. *Scattering Model*

In order to model the scattering process we use the traditional Lambertian model [2], which permits to derive the returned intensity from the parameters defining the observed scene. This simple model for diffuse scattering assumes that the returned intensity depends only on the angle of incidence of the illuminating sound pulse, and not on the angle of observation or on the frequency of the pulse. Under this assumptions the

intensity I returned from a seabed point $\vec{p}$ can be represented by the following expression:

$$I(\vec{p}) = \Phi(\vec{p})\, R(\vec{p}) \cos(\theta(\vec{p})) \tag{1}$$

Where Φ represents the intensity of the illuminating sound wave at point $\vec{p}$, R is the reflectivity of the seafloor, and θ is the incidence angle of the wave front. Since most logged side-scan images already include a Time-Varying Gain (TVG) correction for compensation of the intensity decay with distance and grazing angle, no dependence on radial decay has been included in the model (this would otherwise appear as a term on $r(\vec{p})^{-2}$, r being the distance to the sensor). Therefore, in order to simplify the model, all the intensity variations caused by the sensor's beam-profile, the radial decay and the corrections are supposed to be grouped under the beam-pattern Φ.

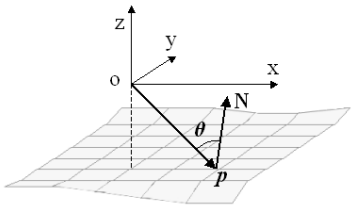

Fig.2: Coordinate system centered on the sensor, at *o*.

The dependence on the seafloor's elevation is implicit in the incidence angle $\theta(\vec{p})$, which depends on the grazing angle from the acoustic source and the orientation of the surface normal $\vec{N}(\vec{p})$. This dependence can be made explicit by first expanding the cosine in expression (1) as follows:

$$\cos(\theta(\vec{p})) = \frac{\vec{r}(\vec{p}) \cdot \vec{N}(\vec{p})}{\left|\vec{r}(\vec{p})\right| \cdot \left|\vec{N}(\vec{p})\right|} \tag{2}$$

And then by representing $\vec{N}$ and $\vec{r}$ on a coordinate system relative to the sensor (Fig.2). Expressing $\vec{p}$ as $(x, y, Z(x, y))$—with x being the across distance from the sensor and y pointing along its direction of movement—gives:

$$\begin{aligned} \vec{r}(\vec{p}) &= (x, 0, Z(x, y)) \\ \vec{N}(\vec{p}) &= \left(-\frac{\partial Z}{\partial x}(x, y), -\frac{\partial Z}{\partial y}(x, y), 1 \right) \end{aligned} \tag{3}$$

Where the gradients $\partial Z/\partial x$ and $\partial Z/\partial y$ can be approximated by finite differences, yielding an expression that depends directly on Z.

IV. PARAMETER ESTIMATION

A. *Expectation-Maximization*

Combination of expressions (1), (2) and (3) gives a formula for the computation of the intensity I at any point $\vec{p}$, given the model parameters R, Z and Φ. The inverse problem is under-determined, since we only have one observation (of I) at each point to compute the values of the three model parameters.

In order to solve this problem we propose to use an expectation-maximization procedure (Fig.3), which will iteratively converge to an optimal set of modeling parameters given a source side-scan image I. The objective is to minimize the absolute value of the difference between the observed intensity I and the one resulting from the application of the model $\hat{I}$, which we represent by the error quantity E:

$$E = \sum_{x,y} E(x, y) = \sum_{x,y} \left(I(x, y) - \hat{I}(x, y) \right)^2 \tag{4}$$

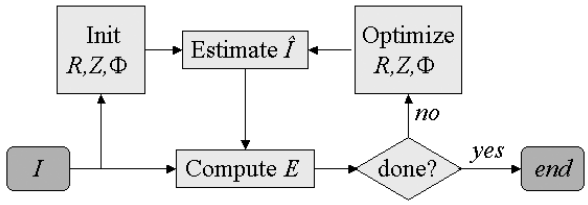

Fig.3: Outline of the Expectation-Maximization procedure.

In the expectation stage, the current estimates for the model parameters are used to compute an estimation of the intensity $\hat{I}(\vec{p})$. This is achieved by substituting the parameters $R(\vec{p})$, $\Phi(\vec{p})$ and $\theta(\vec{p})$ from the previous iteration in the forward model presented in expression (1). Values for $\theta(\vec{p})$ are directly derived from the elevation map Z.

In the maximization stage a straightforward gradient descent approach [11] is used to minimize E, by updating the model parameters as follows:

$$\begin{aligned} R(x, y) &\leftarrow R(x, y) - \lambda \cdot \frac{\partial E}{\partial R}(x, y) \\ \Phi(x, y) &\leftarrow \Phi(x, y) - \lambda \cdot \frac{\partial E}{\partial \Phi}(x, y) \\ Z(x, y) &\leftarrow Z(x, y) - \lambda \cdot \frac{\partial E}{\partial Z}(x, y) \end{aligned} \tag{5}$$

Where λ is a small constant value used to control the rate of change. The expressions are iterated until the variation in the error E is below a given threshold.

B. *Initialization and Regularization*

The optimization procedure starts by initialization of the R, Z and Φ maps. Reflectivity and beam-pattern maps are set to uniform values of 0.9, while the elevation of every point (x, y) is set to that of the first return at $(0, y)$—which is equivalent to the traditional assumption of a flat seabed.

Regularization is performed at the end of every iteration by filtering the reflectivity and beam-pattern maps. Reflectivity values for the points in shadowed areas are set to that of their nearest illuminated neighbors by hexadecagonal dilation [12] of non-shadowed areas. Whereas values of Φ for all the points subtending the same angle to the sensor are set to their median value, since the beam profile of the sensor is supposed to be constant.

Values of R and Φ are normalized to 1. However, because Φ includes the unknown applied TVG, we allow it to achieve values greater than 1. In practice this amounts to just a little overshoot for the bigger angles, which naturally correspond to points of the seabed farther away from the sensor and therefore require higher TVG corrections.

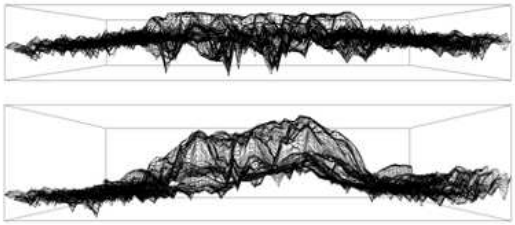

Fig.4: Front-view of the reconstruction of a rock using single-stage (top) and 3-level multi-resolution (bottom) implementations of the proposed reconstruction method. The single stage is not capable of recovering shape features of a bigger scale.

C. *Multi-Resolution*

A multi-resolution implementation of the method described at paragraphs A and B above results in better convergence and improved results. The main reason being the limitation imposed by the point-wise nature of the expectation-maximization procedure, which operates on a per-pixel level. The regularization stage is able to restore some of the interdependence of the pixel neighborhoods, but bigger seabed features, such as slowly varying slopes, cannot be fully recovered. Effects of this limitation are shown in the top part of Fig.4, where the full shape of the rock hasn't been completely recovered on the single-stage implementation of the proposed method.

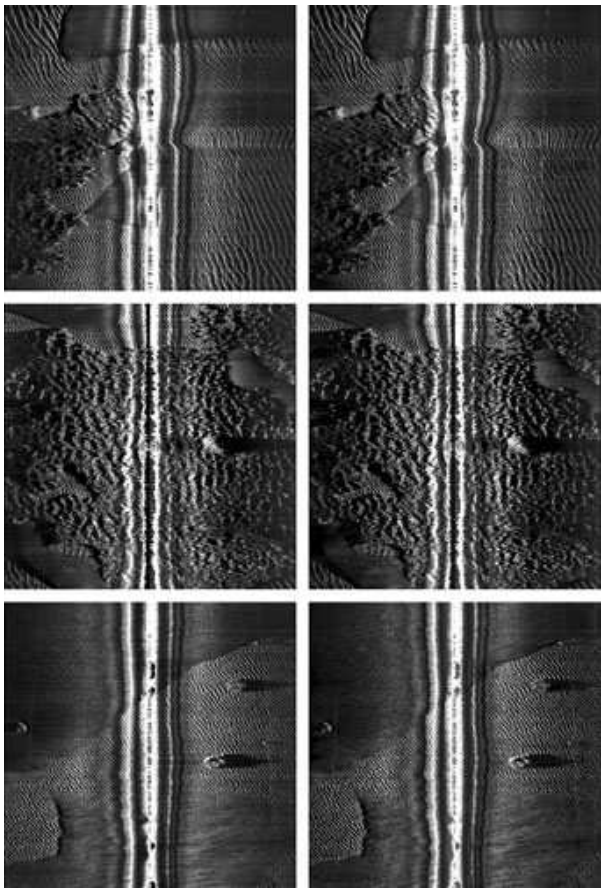

Fig.5: Ground-range images (left column) and synthetic models (right column) after convergence, for three different source side-scan images of the same survey mission.

Results improve notably when using a multi-resolution version of the same algorithm, which is able to recover the seafloor scenes in a more natural way, as well as reducing the overall error at convergence. The bottom part of Fig.4 shows the results of this approach, where the shape of the rock is better estimated, once that more of the spatial frequencies involved are taken into account.

Implementation of the multi-resolution version starts by the construction of a multi-resolution pyramid by iterated sub-sampling of the source side-scan image. Processing starts at the smallest level, using the initialization procedure described in the previous section. The resulting R, Z and Φ maps from one level are used as initial maps for the next resolution level. The process finishes when the final stage—corresponding to the full resolution image—is processed.

V. RESULTS

For testing the consistence of the proposed method, three different images from the same survey, using the same sensor, have been processed. Ground-range corrected side-scan images are shown in the left column of Fig.5. The corresponding synthetic models after convergence of the 3-level multi-resolution implementation are shown in the right column. The ground-range images and the synthetic models are extremely similar, showing the feasibility of the reconstruction approach proposed in this paper.

The full set of outputs of the reconstruction process for the image in the bottom row of Fig.5 is shown in Fig.6. The effect of the beam-profile permeates the reflectivity and elevation maps in the region right under the sensor path, where the Lambertian model cannot properly approximate the non-diffuse reflections. Results are nonetheless consistent with the observed features on the source image. The sand ripples and the structure of the rock are clearly visible in the perspective view of the 3D surface.

The beam-profiles recovered from the three source images are compared in Fig.7. Apart from the inaccuracies corresponding to the region right below the sensor, the shape of the main lobes is consistent across the three profiles, suggesting that, in effect, the same sensor and TVG settings have been used for the acquisition of the three source images.

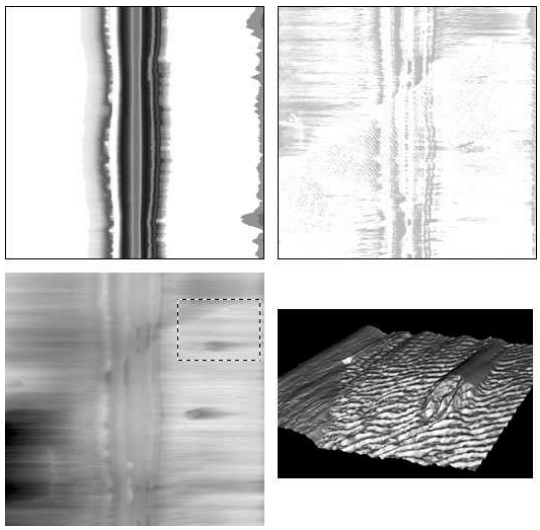

Fig.6: Top to bottom, left to right: beam-pattern, reflectivity and elevation maps; and a perspective view of a 3D surface constructed from the highlighted region in the elevation map.

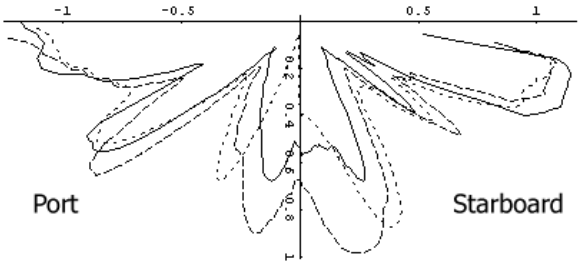

Fig.7:Beam-profiles recovered from the three source images.

Fig.8 shows the evolution of the overall error *E* with the number of iterations for the source image corresponding to the bottom row of Fig.5. The 3 and 5-level multi-resolution versions perform better than the single-stage, due to better initialization.

Finally, a textured version of the 3D render shown in Fig.6 is presented in Fig.9, where the ground-range image has been used as texture.

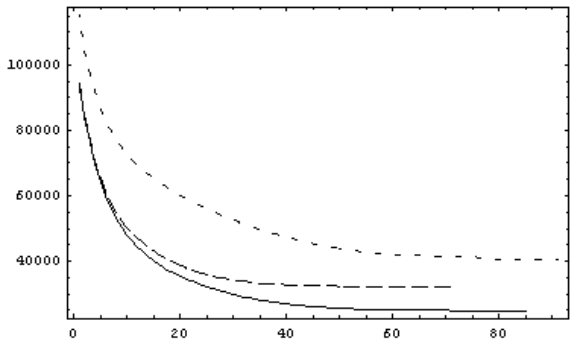

Fig.8: Evolution of the overall error with the number of iterations for: single-stage (short dash), 3-level multi-resolution (long dash) and 5-level multi-resolution (continuous line).

VI. CONCLUSIONS AND FUTURE WORK

In this paper we have presented a new method for the estimation of seabed elevation. The method uses a Lambertian model for the sonar scattering process, which is then used by an expectation-maximization procedure to optimally determine the seabed features ultimately responsible for the observed side-scan image. An example has been presented, which highlights the type of results that can be expected from the proposed method.

For proper rectification of the full seabed elevation and reflectivity maps, however, further work is required. Calibration against ground-truthed scenes needs to be done in order to determine the real accuracy of the method. Finally, transformation from the sensor coordinate frame into a geographical reference system—using the vehicle's navigation data—is required to obtain geo-referenced maps compatible with standard GIS packages.

Applications of the proposed method are numerous, and include accurate mosaic construction, detail improvement on existing bathymetry maps, generation of three-dimensional models of underwater structures, etc.

Acknowledgments

The authors wish to thank the NATO Undersea Research Centre (NURC) for providing the source data used in this paper.

Fig.9: Render of the 3D surface shown in Fig.6, using the corresponding region of the ground-range image as the texture.

REFERENCES

[1] E. Coiras, Y. Petillot, D.M. Lane, "Automatic Rectification of Side-Scan Sonar Images", *Proceedings of the UAM'05 Conference*,Heraklion (Greece), July 2005, in press.

[2] D. Langer, M. Hebert, "Building Qualitative Elevation Maps From Side Scan Sonar Data For Autonomous Underwater Navigation", *Proceedings of the 1991 IEEE International Conference on Robotics and Automation*, vol. 3, pp. 2478 -2483.

[3] R. Li, S. Pai, "Improvement of Bathymetric Data Bases by Shape from Shading Technique using Side-Scan Sonar Images". *Proceedings of the 1991 Oceans Conference*, pp. 320-324, 1991.

[4] S. Tiwari, "Mosaicking of the Ocean Floor in the Presence of Three-Dimensional Occlusions in Visual and Side-Scan Sonar Images", *Proceedings of the 1996 Symposium on Autonomous Underwater Vehicle Technology*, pp. 308 –314, 1996.

[5] J.M. Cuschieri, M. Hebert, "Three-Dimensional Map Generation From Side-Scan Sonar Images", *Journal of Energy Resources Technology - Transactions of the ASME*, 112(2), pp. 96-102, 1990.

[6] A.E. Johnson, M. Hebert, "Seafloor Map Generation for Autonomous Underwater Vehicle Navigation", *Autonomous Robots*, 3(2-3), pp. 145-168, 1996.

[7] V. Murino, A. Trucco, "Three-Dimensional Image Generation and Processing in Underwater Acoustic Vision", *Proceedings of the IEEE*, 88(12), pp. 1903-1948, 2000.

[8] Y.R. Petillot, S.R. Reed and J.M. Bell, "Real Time AUV Pipeline Detection and Tracking Using Side Scan Sonar and Multi-Beam Echosounder", *Proceedings of the MTS/IEEE Oceans Conference*, 2002.

[9] J.M. Bell, E. Dura, S. Reed, Y.R. Petillot, D.M. Lane, "Extraction and Classification of Objects from Sidescan Sonar", *Proceedings of the IEE Workshop on Nonlinear and Non-Gaussian Signal Processing*, 2002.

[10] A.P. Dempster, N.M. Laird, D.B. Rubin, "Maximum Likelihood from incomplete data via the EM algorithm", *J. Royal Statistical Soc., Ser. B*, vol. 39(1), pp. 1-38, 1977.

[11] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, "Numerical Recipes in C. Second Edition", *Cambridge University Press*, p. 421, 1994.

[12] E. Coiras, J. Santamaria, C. Miravet, "Hexadecagonal Region Growing", *Pattern Recognition Letters*, 19, pp. 1111-1117, 1998.