3D Mosaicing for Environment Reconstruction

Vittorio Murino, Andrea Fusiello, Nicola Iuretigh
Dip. Scientifico e Tecnologico
University of Verona, Italy
{murino,fusiello}@sci.univr.it

Enrico Puppo
Dip. di Informatica e Scienze dell'Informazione
University of Genova, Italy
puppo@disi.unige.it

Abstract

This paper proposes a technique for the three-dimensional reconstruction of an underwater environment from multiple range views. The final target of the work lies in improving the understanding of a human operator guiding an underwater remotely operated vehicle (ROV) equipped with an acoustic camera, which provides a sequence of 3D images in real time. Since the field of view is narrow, we devise a technique for the reconstruction of relevant information of the image sequence up to build a mosaic of the surrounding scene. Due to the very noisy nature of the data and the low range resolution, smoothing, segmentation, registration, and fusion problems have been tackled. Examples on real images have been presented to show the promising performances of the algorithm.

1. Introduction

Underwater scene understanding is an expanding field due to the increasing interest in monitoring the evolution of the subsea flora and fauna and the effects of the human interaction with such environment. Moreover, there is also a growing attention in designing new smart sensors able to provide data with a quality impossible to imagine only few years ago. For these reasons, more and more scientists have started to exploit computer vision techniques for underwater image understanding. Our work is a step forward in this direction. We are presenting a technique for three-dimensional (3D) scene reconstruction from a sequence of range data acquired by an acoustic camera. The final goal is to provide a 3D scene model to the human operator(s) of an underwater remotely operated vehicle (ROV), in order to facilitate the navigation and the understanding of the surrounding environment.

Our data come from a high frequency acoustic camera [9]. Speckle noise is typically present due to the coherent nature of the acoustic signals. The noise corrupts sensibly the acoustic signals and decreases the reliability of the estimated 3D measures. Moreover, there is a trade off between range resolution and field of view. Resolution depends on the frequency of the acoustic signal (it is about 3 cm at 500 KHz): the higher the frequency, the higher the resolution, the narrower field of view. Consequently we are forced to operate with a limited field of view and a technique to reconstruct progressively the scene while the sensor is moving is necessary.

In summary, our work aims at reconstructing a 3D environment from a sequence of clutter, noisy, and low resolution data, in order to produce a 3D panoramic mosaic of the scene. This case is quite different from the registration of a couple or more range images, proposed in many previous papers [3, 4, 19]. In fact, we would like to stress that: i) the resolution is never better than some centimeters, unlike classic range data (e.g., from laser range finders); ii) sensor position is not taken into account for view registration; iii) the motion of the sensor is quite unstable, and cannot be controlled with precision in any real case, so acquired images from a fixed position may be different due to speckle and sensor floating. As a consequence, some previous solutions based on estimation of surface parameters cannot be taken into account due to the high uncertainty of the data. A lot of works on the registration of image pairs or the integration of a set of range images are present in literature, but none dealing with the particular kind of 3D data we are using. So, although the problems we encountered may seem the same discussed in other papers, significant differences are actually found.

Among the works related to registration, the Iterative Closest Point (ICP) procedure [3] and its earlier variants [4, 19] are seminal papers worth to be mentioned.

The work in [4] also deals with the possibility to register more range images by incrementing pairwise registration, resulting in an non optimal global registration. Other works address this problem [2, 1, 15, 13]. As an example, in [2], couples of images are incrementally registered together with a final registration between the first and last view, by using the inverse calibration procedure of the range-finder to relate a point in the 3D space corresponding to a point in the range image. Another work [1] considers all the views simultaneously and tries to minimize the global registration error, assuming that the error over the several views is little (i.e., a preliminary rough registration has been applied). The work in [16] aims at estimating surface approximations from several range images, and subsequently integrating them in a global surface model. No constraints about the

1051-4651/00 $10.00

number of views, viewpoints' positions, and object topology are to be respected. In [15] and [6], a force-based optimization approach is proposed. Assuming the points' correspondences among the data sets known, interconnections using springs between corresponding points is simulated. More recently, a multiview registration technique has been presented in [13]. The method starts using a pairwise registrations (ICP-based) between closest views, and use these results as constraints for the multiview alignment. In such a way, computational time is reduced as well as memory storage, and pairwise registration error is spread among the views rather than accumulating.

All these works assume range images from a laser range finder looking at a single, even complex, object. Unlike these ones, our algorithm deals with uncertain low resolution data, in which problems of filtering, segmentation and reconstruction should be all considered in order to get a reliable reconstruction of the scene.

In our approach, we first pre-process rough data in order to reduce noise, eliminate clutters, and group together points belonging to different surface patches (Sec. 2). Then, resulting data patches are smoothed using a robust fitting technique (Sec. 3). Finally, individual range images are pairwise registered, using a variation of ICP, and then fused together in order to generate a mosaic containing all the available data (Sec. 4). Examples on real images of an underwater oil rig are presented (Sec. 5) showing the promising performances of the algorithm.

2. Pre-processing

Three-dimensional data are obtained by a high resolution acoustic camera, the *Echoscope* [9]. The scene is insonified by a high-frequency acoustic pulse and a two-dimensional array of transducer gathers the backscattered signals. The whole set of raw signals is then processed in order to enhance those coming from fixed steering directions (called *beamsignals*) and to attenuate those coming from other directions. The distance of a 3-D point can be measured by detecting the time instant at which the maximum peak occurs in the beamsignal. A range image is formed by 64×64 points ordered according to an angular relation, as adjacent points correspond to adjacent beamsignals. Moreover, the intensity of the maximum peak can be used to generate another image, representing the reliability of the associated 3-D measures: the higher the intensity, the safer the associated measure.

The acoustic image is affected by false reflections, caused by secondary lobes, and by acquisition noise, which is modeled as speckle noise. The intensity image turns out to provide very useful information to discriminate between "good data" and noise. A dramatic improvement of the range image quality is obtained by discarding points whose associated intensity is lower than a threshold depending on

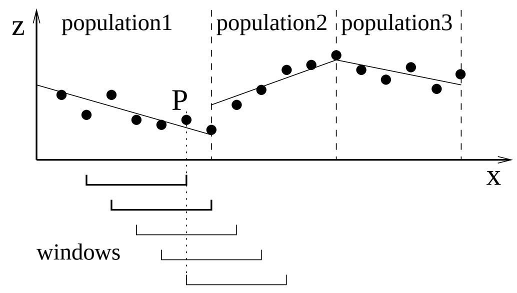

Fig. 1. Pixel P belongs to five different windows; the first two leftmost windows crosses a homogeneous population.

the secondary lobes. Then, a size filter is applied considering as connected range points closer than a given threshold.

The cleaned range image is then converted from polar sensor coordinates (ρ, θ, r) to Cartesian world coordinates (x, y, z). The benefit of converting to Cartesian coordinates lies in the ability to fuse range data collected from different viewpoints into a single representation.

Then a surface mesh is created based on a Delaunay triangulation on the (x, y) plane. To group points into surface patches, edges longer than a specified threshold are removed, leaving the corresponding nodes unconnected. This surface mesh is then resampled on a rectangular grid, by fitting a plane on each triangle.

The resulting resampled image constitute a range image in Cartesian space which is used in all subsequent processing stages. However, the data must be further processed to remove invalid data points (outliers) and sensor noise, while preserving range discontinuities. To this end a spatial smoothing filter is applied, as described in the next section.

3. Smoothing by line fitting

Smoothing of range data is usually performed by locally fitting a parametric surface to range data. However, many methods based on such technique produce inaccurate results when surface or derivative discontinuities are present, and even on smooth surfaces, whenever the image contains scattered impulse values, called *outliers*. Thus, *robust* methods are needed, that are powerful enough to handle data coming from discontinuous (piecewise-smooth) surfaces and affected by different kinds of noise. When two-dimensional windows are used, points within a window close to surface discontinuities come from at least two different populations: "drawing the line" between the two populations is in general a difficult problem. When more than two populations are considered (e.g., close to object corners) or non-straight edges cross the window, the problem is even harder. In order to overcome this problem, following [14], we use one-dimensional (linear) windows swept along several directions on the image plane, and integrate the results ob-

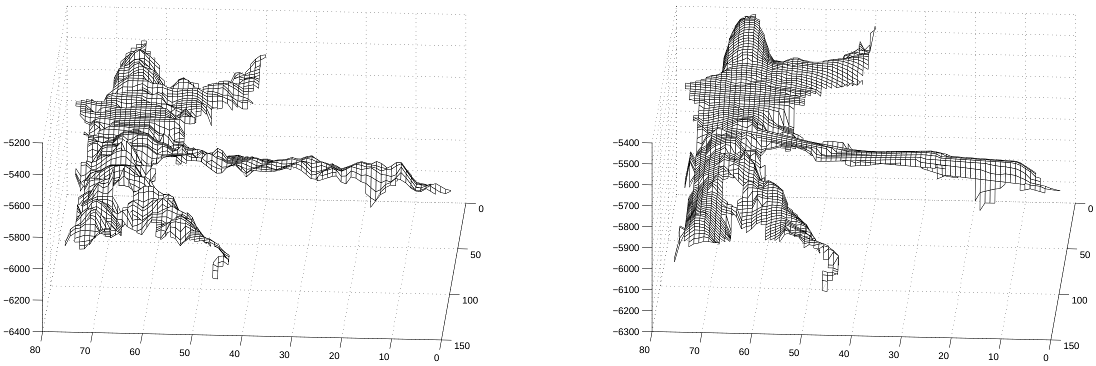

Fig. 2. Acoustical data before (left) and after (right) smoothing.

tained through this first *directional processing* step for obtaining a solution to the two-dimensional problem. This algorithm solves the surface fitting problem on range data in a fast, highly-parallel, efficient yet simple and robust way. Although it is suited for the special case of piecewise-linear surfaces, the method can be generalized to polynomial surfaces of higher degree.

The method consists of two distinct and independent steps. First, an isotropic set of directions is taken, and *slices* are extracted from the original image, in such a way that every pixel belongs to exactly one slice per direction. Each slice is then viewed as the discrete, noisy version of a piecewise-smooth function of one variable. A one-dimensional fitting algorithm is applied to each slice; the one-dimensional processing for every direction gives an estimate for the z value at each point. In the second step, all estimates obtained through the one-dimensional algorithm are considered. Different estimates of the position of one pixel are averaged in order to obtain a final estimate.

As for the first step, the method locally solves the problem by allowing the neighborhood of each pixel to "float around", looking for a homogeneous set of data, i.e. a set which does not contain any discontinuity. The method is based on the assumption that at least one partial neighborhood per pixel always exists whose data are homogeneous (see Fig, 1). For every point P, the algorithm fits one line per window on all windows containing P, basing on Haralick's facet model [10]. A *goodness-of-fit* measure is computed for each window, then the "best" result is chosen to give estimates for the value of the underlying function at point P. We have experimented with various algorithms and goodness-of-fit measures. Under mild assumptions on the nature of the noise, a least-square algorithm and a χ^2 measure can be successfully applied, provided that data are pre-processed. This solution achieves a good compromise between accuracy and computational complexity. An example of smoothing on a real underwater acoustic image is shown in Fig. 2.

4. Registration and Fusion

Building of global model from a sequence of unregistered range images, is achieved in two stages: (i) pairwise registration of each subsequent frame to compute the rigid displacement that transform measurements into a common coordinate frame; (ii) fusion of all the sets of measurements into a single 3D surface. Registration was addressed using the classical Iterative Closest Point (ICP) algorithm [3], a general purpose method for the registration of rigid 3-D shapes.

4.1 ICP based algorithm

Let us suppose that we have two sets of 3-D points which correspond to a single shape but are expressed in different reference frames. We will call one of these sets the model set X, and the other the data set Y. Let us start by assuming that for each point in the data set, the corresponding point in the model set is known. The problem is to find a 3-D transformation which, when applied to the data set Y, minimizes a distance measure between the two point sets. The goal of this problem can be stated more formally as follows:

$$\min_{\mathbf{R},\mathbf{t}} \sum_{i=1}^{N} \|\mathbf{x}_i - (\mathbf{R}\mathbf{y}_i + \mathbf{t})\|^2, \qquad (1)$$

where $\mathbf{R}$ is a 3×3 rotation matrix, $\mathbf{t}$ is a 3×1 translation vector, and the subscript i refers to corresponding elements of the sets X and Y. Efficient, non-iterative solutions to this problem were compared in [11], and the one based on Singular Value Decomposition (SVD) was found to be the best. In general, however, point correspondences are unknown. For each point $\mathbf{y}_i$ from the set Y, there exists at least one point on the surface of X which is closer to $\mathbf{y}_i$ than all other points in X. This is the *closest point*, $\mathbf{x}_i$. The basic idea behind the ICP algorithm is that, under certain conditions, the point correspondence provided by sets of closest points is a reasonable approximation to the true point correspondence. Besl and McKay [3] proved that if the process of finding closest point sets and then solving Eq. (1) is repeated, the solution is guaranteed to converge to a local minimum. The ICP algorithm can be summarized:

1. For each point in Y, compute the closest point in X;
2. With the correspondence from step 1, compute the incremental transformation $(\mathbf{R}, \mathbf{t})$ with SVD;
3. Apply the incremental transformation from step 2 to the data Y;
4. If the change in total mean square error is less than a threshold, terminate. Else goto step 1.

Modifications to the original ICP are now widely used to achieve accurate registration of pairs of *partially* overlapping range images [19, 18, 17]. We implemented a variation similar to the one proposed by Zhang [19], using a modified cost function based on robust statistics to limit the maximum distance between closest points. As pointed out by Zhang, the distribution of the residuals for two fully overlapping sets approximates a Gaussian, when the registration is good. The non-overlapped points skew the distribution of the residuals, hence the threshold on the distance must be set using a robust statistics. Following the X84 rule [8] we discard those points whose residual differ more than 5.2 MAD (Median Absolute Deviations) from the median. The value 5.2 corresponds to about 3.5 standard deviations, which encloses more than 99.9% of a Gaussian distribution.

As it is well known, pairwise registration of pairs of overlapping point sets does not yield the optimal result. Many method have been presented to optimally solve this problem (see the literature review in Sec. 1). Yet, using a small number of views, accuracy was satisfactory for our purposes. Severe time constraints did not allow us to take global refinement into consideration.

4.2 Fusion

Registered sets of points must be fused in order to get a single 3D model. After registering frame i with frame $i-1$, the former is fused with the current model, built from frames $1 \ldots i-1$. Fusion occurs between 3D points: every point of the current frame which is closer to a point of the model less than a threshold is deleted. In this way, every subsequent frame add a portion to the model. We avoided both mesh integration [18, 16] and volumetric fusion [5, 12] because of the high computational load.

5. Results

As an example of the results obtained, we report in Fig. 3 frames number 1,4,7, and 10 of a sequence of ten range images, taken by the Echoscope mounted on a underwater ROV. The subject is a joint where five pipes meet, which is part of a off-shore platform. These frames are the output of the pre-processing and smoothing stages described in Sec. 2 and Sec. 3. Fig. 4(a) shows the final mosaic of the sequence, computed as described in Sec. 4. Fig. 4(b) shows the mosaiced data along with cylinders that have been automatically fitted using a method that we developed [7].

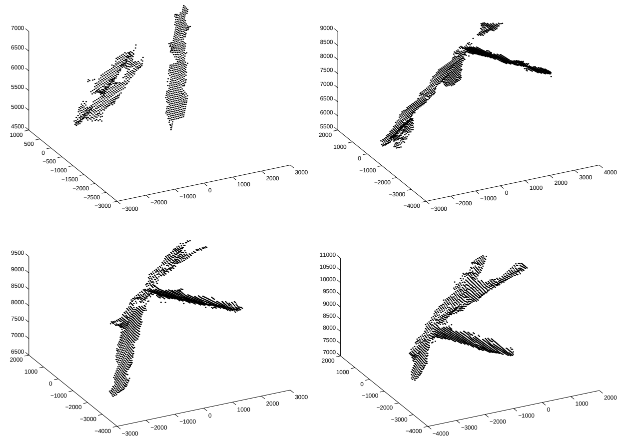

Fig. 3. Samples from a ten-frames sequence of range images.

6. Conclusions

In this paper, we presented an ICP-based approach to the construction of 3D mosaics from acoustic range data. After a coordinate transformation and noise/clutter filtering phase, dense range images are generated and smoothed to limit the problem of coarse accuracy and range reliability. Then, images are pairwise registered and incrementally fused together.

Although the application scenario is constituted by pipe-like structures, the proposed method is completely general and no constraints are given on the number of views, resolution, and viewpoints' locations, provided that a certain degree of overlapping is present.

Improvements in the registration of multiple views are foreseen in order to achieve a finer accuracy, possibly better than the sensor resolution. As for the fusion, our effort will be aimed at investigating a method that smooth out the noise while fusing, by taking into account all the data, not just the latest.

Methods for the automatic setting of the thresholds will be also considered.

The amount of overlapping necessary for a good registration is left unspecified, but, from an applicative point of view, this is not a problem as, in underwater environments, the vehicle is slowly moving and a sufficiently large overlapping is always present.

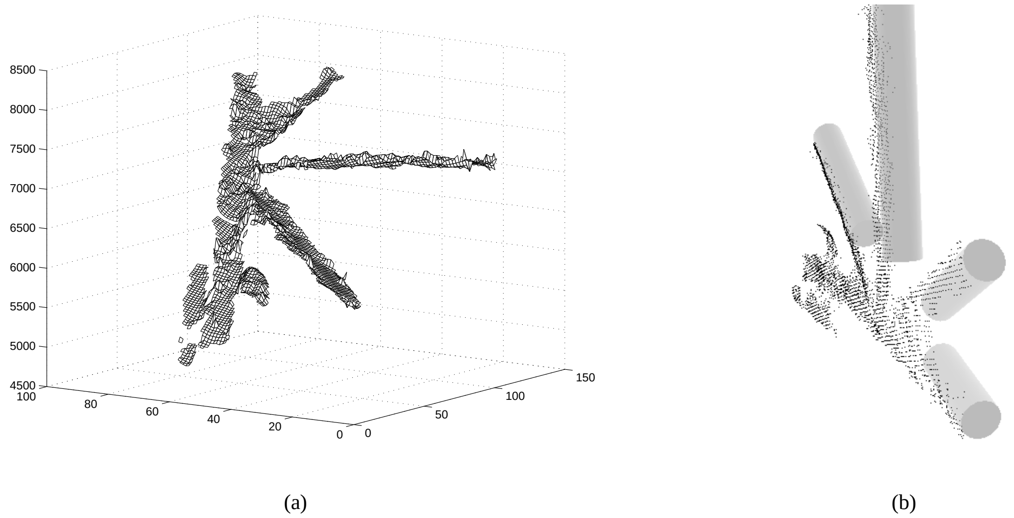

Fig. 4. (a): mosaic of the sequence of Fig. 3. (b): mosaic with fitted cylinders superimposed.

Acknowledgements. This work was supported by the European Commission under the BRITE-EURAM III project no. BE-2013 VENICE (Virtual Environment Interface by Sensor Integration for Inspection and Manipulation Control in Multifunctional Underwater Vehicles). Thanks to R. Giannitrapani for fruitful discussions, and to Dr. R.K. Hansen of Omnitech A/S for providing the Echoscope images.

References

[1] R. Bergevin, M. Soucy, H. Gagnon, and D. Laurendeau. Towards a general multiview registration technique. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(5):540–547, May 1996.

[2] G. Blais, M.D. Levine. Registering Multiview Range Data to Create 3D Computer Objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(8):540-547, 1995.

[3] P. Besl and N. McKay. A method for registration of 3-D shapes. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(2):239–256, February 1992.

[4] T. Chen and G. Medioni.. Object modeling by registration of multiple range images. *Image and Vision Computing*, 10(3):145–155, 1992.

[5] B. Curless, M. Levoy. Volumetric Method for Building Complex Models from Range Images. In Computer Graphics Proceedings, *SIGGRAPH '96*, 1996.

[6] D. W. Eggert, A. W. Fitzgibbon, and R. B. Fisher. Simultaneous registration of multiple range views for use in reverse engineering. In *Proceedings of the Int. Conference on Pattern Recognition*, pp. 243–247, Vienna, 1996.

[7] A. Fusiello, R. Giannitrapani, V. Isaia, and V. Murino. Virtual environment modeling by integrated optical and acoustic sensing. In *Int.Conf. on 3D Digital Imaging and Modeling*, pp 437-446, Ottawa, 1999

[8] F. Hampel, P. Rousseeuw, E. Ronchetti, and W. Stahel. *Robust Statistics: the Approach Based on Influence Functions*. Wiley Series in probability and mathematical statistics. John Wiley & Sons, 1986.

[9] R. K. Hansen and P. A. Andersen. A 3-D underwater acoustic camera - properties and applications. In P.Tortoli and L.Masotti, editors, *Acoustical Imaging*, pp 607–611. Plenum Press, 1996.

[10] R. Haralick and L. Watson. A facet model for image data. *Computer Graphics Image Processing*, 15:113–129, 1981.

[11] A. Lorusso, D. W. Eggert, and R. B. Fisher. A comparison of four algorithms for estimating 3-D rigid transformations. *Machine Vision and Applications*, 9:272–290, 1997.

[12] K. Pulli, T. Duchamp, H. Hoppe, et al. Robust Meshes from Multiple Range Maps. In *Int. Conf.on 3D Digital Imaging and Modeling '97*, Ottawa, pp. 205-211, 1997.

[13] K. Pulli. Multiview Registration for Large Data Sets. *Int.Conf. on 3D Digital Imaging and Modeling*, Ottawa, pp.160-168, 1999.

[14] E. Puppo and L. Davis. Surface fitting of range images using a directional approach. Technical Report CAR-TR-504, Center for Automation Research, Univ. of Maryland, 1990.

[15] A.J. Stoddart, A. Hilton. Registration of Multiple Point Sets. In *Int. Conf. on Pattern Recognition*, Vienna, pp.40-44, 1996.

[16] M. Soucy, D. Laurendeau. A General Surface Approach to the Integration of a Set of Range Views. *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 17(4):344-358, April 1995.

[17] E. Trucco, A. Fusiello, and V. Roberto. Robust motion and correspondences of noisy 3-D point sets with missing data. *Pattern Recognition Letters*, 20(9):889-898, September 1999.

[18] G. Turk and M. Levoy. Zippered polygon meshes from range images. In *SIGGRAPH 94*, pp. 311–318, Orlando, 1994.

[19] Z. Zhang. Iterative point matching of free-form curves and surfaces. *Int. Journal of Computer Vision*, 13(2):119–152, 1994.